

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:36:32 ; Search time 87.3574 Seconds
(without alignments)
13131.604 Million cell updates/sec

Title: US-09-602-362E-15
Perfect score: 3403
Sequence: 1 ctgcgcccgttaagatggct.....gaaataaaaaaanaaaaaa 2030

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV=xlp
-MODE=frame+n2p.model -DEV=xlp
-O=/gn2_1/USPTC spool_p/US0602362/runat_15072004_093625_21981/app_query.fasta_1.10325
-DB=A_Geneseq_25Jan04 -QMT=fascan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @CGN 1.1.729 @runat 15072004_093625_21981 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOF=6
-FGAPEXT=7

Database : A_Geneseq_25Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3272	95.2	1341	4 AAB84702	Aab84702 Amino aci
2	3272	95.2	1341	5 ABJ05537	Abj05537 Breast ca
3	3272	95.2	1341	6 ABJ37784	Abj37784 Human tum
4	3272	95.2	1341	6 ABR47548	Abr47548 Breast ca
5	3272	95.2	1349	6 ABJ37788	Abj37788 Human tum
6	3261	95.8	1002	4 AAU33351	Aau33351 Human bre
7	3261	95.8	1002	5 ABG78918	Abg78918 Human bre
8	3261	95.8	1002	6 ABJ37741	Abj37741 Human tum
9	3261	95.8	1002	8 ADE44427	Ade44427 Human bre
10	3261	95.8	1095	4 AAU33357	Aau33357 Human bre

11	3261	95.8	1095	5 ABG78924	Abg78924 Human bre
12	3261	95.8	1095	6 ABJ37747	Abj37747 Human tum
13	3255	95.7	1013	6 ABJ37783	Abj37783 Human tum
14	2599	76.4	512	4 AAB84701	Aab84701 Amino aci
15	2256	66.3	1239	6 ABJ37789	Abj37789 Human tum
16	2184	64.2	445	4 AAB50249	Aab50249 Human bre
17	2184	64.2	445	4 AAG55987	Aag55987 B726P spl
18	2184	64.2	445	4 AAU33350	Aau33350 Human bre
19	2184	64.2	445	5 ABG78917	Abg78917 Human bre
20	2184	64.2	445	6 ABJ37740	Abj37740 Human tum
21	2184	64.2	445	8 ADE44425	Ade44425 Human bre
22	2017	59.3	432	4 AAB50244	Aab50244 Human bre
23	2017	59.3	432	4 AAG55982	Aag55982 B726P thi
24	2017	59.3	432	4 AAU33345	Aau33345 Human bre
25	2017	59.3	432	5 AAO15779	Aao15779 Human bre
26	2017	59.3	432	5 ABG78912	Abg78912 Human bre
27	2017	59.3	432	6 ABJ37735	Abj37735 Human tum
28	2017	59.3	432	8 ADE44133	Ade44133 Human bre
29	1785	52.5	1225	6 ABR47547	Abr47547 Breast ca
30	1597.5	46.9	650	4 AAB50263	Aab50263 Human bre
31	1597.5	46.9	650	4 AAG55983	Aag55983 B726P spl
32	1597.5	46.9	650	4 AAU33346	Aau33346 Human bre
33	1597.5	46.9	650	5 ABG78913	Abg78913 Human bre
34	1597.5	46.9	650	6 ABJ37736	Abj37736 Human tum
35	1597.5	46.9	650	8 ADE44421	Ade44421 Human bre
36	1597.5	46.9	743	4 AAU33358	Aau33358 Human bre
37	1597.5	46.9	743	5 ABG78925	Abg78925 Human bre
38	1597.5	46.9	743	6 ABJ37748	Abj37748 Human tum
39	1594.5	46.9	398	3 AAB07638	Aab07638 Amino aci
40	1591.5	46.8	661	6 ABJ37782	Abj37782 Human tum
41	1546	45.4	1011	4 AAB84703	Aab84703 Amino aci
42	1514	44.5	335	4 AAB63900	Aab63900 Human pro
43	1404.5	41.3	410	4 AAU33359	Aau33359 Human bre
44	1404.5	41.3	410	5 ABG78926	Abg78926 Human bre
45	1404.5	41.3	410	6 ABJ37749	Abj37749 Human tum

ALIGNMENTS

RESULT 1	
AAB84702	
ID	AAB84702 standard; protein; 1341 AA.
XX	XX
AC	AAB84702;
XX	XX
DT	17-SEP-2001 (first entry)
XX	XX
DE	Amino acid sequence of a human cancer associated antigen.
XX	XX
KW	Cancer associated antigen; INGI1; tumour suppressor; cancer; vaccine.
XX	XX
OS	Homo sapiens.
XX	XX
PN	W0200147959-A2.
XX	XX
PD	05-JUL-2001.
XX	XX
PF	29-NOV-2000; 2000WO-US042334.
XX	XX
PR	30-NOV-1999; 99US-00451739.
PR	24-OCT-2000; 2000US-00602362.
XX	XX
PA	(LUDW-) LUDWIG INST CANCER RES.
PA	(SLOK) SLOAN KETTERING INST CANCER RES.
XX	(CORR) CORNELL RES FOUND INC.
XX	XX
PI	Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;
XX	XX
DR	WPI; 2001-441706/47.
XX	XX
PT	Isolated cancer associated nucleic acid molecule identified by SEREX
PT	(serological identification of antigens by recombinant expression
PT	cloning) technique, useful in nucleic acid based therapies to treat

PT cancer.
 XX Claim 83; Page 53-57; 62pp; English.
 XX The present sequence represents a human cancer associated antigen. The INGI
 CC sequence was identified using probes derived from the INGI gene. The INGI
 CC gene is a tumour suppressor candidate gene. The cancer associated antigen
 CC polynucleotides and polypeptides are useful for screening for the
 CC possible presence of a pathological condition in a subject such as
 CC cancer. The cancer associated antigen polypeptides are useful for
 CC producing vaccines
 XX
 XX Sequence 1341 AA;
 SQ

Alignment Scores:
 Pred. No.: 2,04e-274 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 4 Gaps: 0

US-09-602-362E-15 (1-2030) x AAB84702 (1-1341)

QY	9	GTAAAGATGGCTCTCTGAAGCTAACTCGGAATGAAGATTTCTATTCCACTAAAGCC	68
DB	666	ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla	685
QY	69	TTAGAATTGATGACATGCAAACTTTCAAGACAGAGCCCTCCGAGAAGCCATCTGCCTTC	128
DB	686	LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe	705
QY	129	GACCTCCCATGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATTGAAGATGAA	188
DB	706	GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu	725
QY	189	CRAACATGACGACATGAGTACTCCATCAGAAATCCAAACAAAGGACTATCAAGAA	248
DB	726	GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu	745
QY	249	AGTTCTTGGGATCTCAGAGTCTCTGAGACTGTTTTCACAGAGATGTGTTTACCC	308
DB	746	SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro	765
QY	309	AGGCTACACATCAAAAGAAATAGATAAATAATGGAATTAAGAGTCTCTCTGAT	368
DB	766	LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp	785
QY	369	AATGATGGTCTTCTGAAGGCTCCCTCGAGAATGAAAGTTTCTATTCCAACTAAAGCCTTA	428
DB	786	AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu	805
QY	429	GAATGATGACATGCAAACTTTCAAGACAGAGCCCTCCGAGAGCCATCTGCCTCGAG	488
DB	806	GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu	825
QY	489	CCTGCCATTCGAATGCAAAAGTCTGTTCCAAATGAAAGCCCTTGAATTAAGAAATGAACAA	548
DB	826	ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln	845
QY	549	ACATTGAGACGATCAGATGTTCCCTTCGAATCAAAACAAAGAGGTTGAAGAAAT	608
DB	846	ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn	865
QY	609	TCTTGGATTCAGAGTCTCCGTGAGACTGTTTTCACAGAAGGATGTGTGTGTTACCCCAAG	668
DB	866	SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys	885
QY	669	GCTACACATCAAAAGAAATGGAATAAATGAAGTGGAAATTAAGAAATTAAGAAATTAAGAA	728
DB	886	AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu	905
QY	729	TCAAAATCTTGGATACAGTTCATCTTGTGAAAGAGCAAGGAACTTCAAAAGATCAC	788

DB	906	SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis	925
QY	789	TGTGAACAACCTACAGAAAAATGGAACAAATCAAAAGAAAGTTTGTCTACTGAAAAAG	848
DB	926	CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys	945
QY	849	AAACTCTGAGAACCAAAAGAAATAAATCACTAGTAGAACAACAAAGTTAAATGGGAA	908
DB	946	LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu	965
QY	909	CAAGAGCTCTGAGTCTGAGATTGACTTTAAACCAAGAAAGAGAGAGAGAAATGCC	968
DB	966	GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla	985
QY	969	GATATATTAATGAAAAATTAGGGAAGATATTAGGAAGAAATCGAAGACGACGATAGAAA	1028
DB	986	AspIleLeuAsnGlnLysIleArgGluGluLeuGlyArgIleGluGlnHisArgLys	1005
QY	1029	GAGTTAGAGTGAACAACAACCTTGAACAGGCTCTCAGAAATCAAGATATAGAAATGAAG	1088
DB	1006	GluLeuGluValLysGlnGlnLeuGluAlaLeuArgIleGlnAspIleGluLeuLys	1025
QY	1089	AGTGTAGAAAGTAAATTTGAATCAAGTTCTCACATCATGAAATGAAATATCTCTTA	1148
DB	1026	SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu	1045
QY	1149	CATGAAATTCATCTTGAAGAAAGAAATGCGATCTGCTAAACTGGAATAGCACACATG	1208
DB	1046	HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysGluIleAlaThrLeu	1065
QY	1209	AAACACCAATACACAGAAAGAAATTAATATCTTTGAGGACATTAAGATTTTAAAGAA	1268
DB	1066	LysHisGlnTyrGlnGlnLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu	1085
QY	1269	AAGAACTCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAATCATTAATAAGGGA	1328
DB	1086	LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla	1105
QY	1329	TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACACAATGTCTCACTTAAA	1388
DB	1106	SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys	1125
QY	1389	TTGAAGGAAAAACAAGAAAAATCTAGAGGCGAATTAAGATCAACACCATCTCTAGA	1448
DB	1126	LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg	1145
QY	1449	CTGCTTCTCTGTACAGACCATGATCAAAATGTGACATCAAGAAAAAGTCAAGAACCT	1508
DB	1146	LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro	1165
QY	1509	GCTTTCACACATTCGAGGAGATGCTTGTTCGAAAGAAAAATGAATGTTGATGTAGTAGT	1568
DB	1166	AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer	1185
QY	1569	ACCATATATACAAATCAGGTGCTCCATCAACACCTTCTGAAAGCTCAAGGAAATCCAN	1628
DB	1186	Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys	1205
QY	1629	AAGCTTAAAAATTAATCTCAATATTCGAGGAGATGCTTAAAGAGAAAAATACATGTTTC	1688
DB	1205	sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe	1225
QY	1689	AGGAACATGCAAAAGACCAACAGTGAACACAGCTGTCAAATGCAAGGAGCTGAACACA	1748
DB	1225	r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis	1245
QY	1749	TGTTTCAAAACGAACAAGATNATGTGAACAAACACACTGANCAGCAGGAGCTCTTAGATC	1808
DB	1245	etTyrGlnAsnGluGlnAsnValAsnLysHisThrGluGlnGlnLysSerLeuAspG	1265
QY	1809	AGAAATATTTCAACTCAAGGAAAAATATGTGGCTTCAACAGCAATAGTTCATCATCAC	1868

Db 1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAGCTGCAACAACAAAGCAAGATAACAATTGATNTTCATTNTCTGAGAGGAAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTCTTAAAGAGAGAAAATGAGGAGATATTNATTACNATACCATTTA 1987
Db 1305 MetGlnHisLeuLeuLeuLysGlnLysAsnGluGluIlePheAsnTyAsnAsnHisLeu 1324
QY 1988 AAAACCCGCTATATTCATATGGAAGAAAAAANAANA 2027
Db 1325 LysAsn-ArgIleTyGlnTyGluLysGlnLysAlaGlu 1337
RESULT 2
ABC05537
ID ABJ05537 standard; protein; 1341 AA.
AC ABJ05537;
XX
DT 14-NOV-2002 (first entry)
XX
DE Breast cancer-associated protein 2.
XX
KW Breast cancer; breast cancer-associated gene sequence; drug development;
KW pharmacogenetics; biosensor development.
XX
OS Unidentified.
XX
XX WO200259377-A2.
PN
XX
PD 01-AUG-2002.
XX
PF 24-JAN-2002; 2002WO-US002242.
XX
PR 24-JAN-2001; 2001US-0263965P.
PR 02-FEB-2001; 2001US-0265928P.
PR 09-APR-2001; 2001US-00829472.
PR 09-APR-2001; 2001US-0282698P.
PR 04-MAY-2001; 2001US-0288590P.
PR 29-MAY-2001; 2001US-0294433P.
XX
PA (EOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Mack DH, Gish KC, Afar D;
PI
XX
XX WPI; 2002-583738/62.
DR
XX N-PSDB; ABT07694.
XX
PT Detecting a breast cancer-associated transcript in a patient's cell,
PT useful for diagnosing breast cancer, comprises contacting a biological
PT sample with a polynucleotide that selectively hybridizes with breast
PT cancer nucleic acids.
XX
PS Disclosure; Page 348-349; 414pp; English.
XX
XX The invention comprises a method of detecting a breast cancer-associated
CC transcript in a cell from a patient. The method of the invention involves
CC contacting a biological sample from the patient with a nucleotide that
CC hybridizes to one of the 69 breast cancer-associated gene sequences shown
CC in the specification. The method of the invention is useful in the
CC diagnosis or prognosis of breast cancer, and for detecting genes that are
CC up or down-regulated in breast cancer cells. Genes identified by the
CC method of the invention can be used in diagnostic purposes and also as
CC targets for screening for therapeutic compounds that modulate breast
CC cancer (e.g. hormones or antibodies). Identification of genes that are
CC over or under expressed in breast cancer can additionally provide high-
CC resolution, high-sensitivity datasets which can be used in the areas of
CC diagnostics, therapeutics, drug development, pharmacogenetics, protein
CC structure and biosensor development. Amino acid sequences ABJ0536 -
CC ABJ05604 represent the proteins encoded by the 69 breast cancer-
CC associated genes of the invention
XX

SQ Sequence 1341 AA;
Alignment Scores:
Pred. No.: 2,04e-274 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: Gaps: 0
US-09-602-362E-15 (1-2030) x ABJ05537 (1-1341)
QY 9 GTTAAAGATGCTTCTTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCTCAACTAAAGCC 68
Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTAGAATTGATGACATGCAAACTTTCAAGACAGAGCCTCCCGAGAGCCATCTGCCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GAGCTGCCATTGAAATGCAAAAGTCTGTCCAAATAAAGCCTTGCAATTGAAGAATGAA 188
Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAAACATTGAGACAGATGATATCTCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyArgGlu 745
QY 249 AGTCTCTGGATTCTGAGATCTCTGTGAGACTGTTTCACAGAGGATCTGTCTTACCC 308
Db 746 SerSerTrpAspSerGluSerLysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGGTACACATCAAAAGAAATAGATAAAATAAATAGAAATAGAAAGTCTCTCGAT 368
Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
QY 369 AATGATGGTTTCTGAAGGCTCCTCGAGATCAAAAGTTTCTATTCTCAACTAAAGCCTTA 428
Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATTGATGACATGCAAACTTTCAAAGACAGCCTCCCGAGAGCCATCTGCCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCTTGCAATTGAAGAATGAACAA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACAGATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTTGGATTCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGGATCTGTGTGTACCCAAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGGAATAAGTGGAAATAGCAAGATTCACATGACCTA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGATACAGTTTCATTCTTGTGAAGAGCAAGGAACTTCAAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAGCTACAGAAATAATGGAACAAATGGAATAAGAGTTTGTGTACTGAAAGAG 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAATCTGCAGACCAAGAAATAATAATCACACTTAGAGAACCAAAAGTTTAATGGAA 908
Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGlnLysValLysTrpGlu 965
QY 909 CAAGAGCTCTGCAGTGTGAGATTGATTTTAAACCAAGAGAGAGAGAGAAATGCC 968

QY 189 CAACATTGAGACGATGAGATACCTCCATCAGAAATCAACAAAGAGACTATGAGAA 248
 Db 726 GlnThrLeuArgAlaAspGluLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTTCTTGGGATCTGAGAGTCTCTGAGAGTCTTTCACAGAGGATGCTGTTACCC 308
 Db 746 SerSerTrpAspSerGluSerLysGlnThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGGCTACACATCAAAAGAAATAGATAAAATGAAATGAAAGAGTCTCTCTGAT 368
 Db 766 LysAlaThrHisGlnLysGluLeuAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGTTTCTGAGAGCTCCCTGCGAGATGAAAGTTTCTATTCACACTAAGCCCTTA 428
 Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATGTGAGCATCAAACTTTCAAGCAGAGCTCCCGAAGCCATCTGCTTCGAG 488
 Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CTGTCATTGAAATGCAAAAGTCTTTCCTTCCAAATAAAGCTTGGAAATGGAAGTGAACAA 548
 Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACGATCAGATGTTTCCCTTCAGATCABAACAAAGAGGTTGAAGAAAT 608
 Db 846 ThrLeuA-gAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 865
 QY 609 TCTTGGGATCTGAGAGTCTCCGTCGAGACTGTTTTCACAGAGGATGTTGTGTACCCCAAG 668
 Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACATCAAAAGAAATGATGATAAATAGTGAATAATAGAGATTCACATGACCTA 728
 Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
 QY 729 TCAAAATCTTGGATACAGTTCATTCTTGTGAAGCAAGCAAGCAACTTCAAAAGATCAC 788
 Db 906 SerLysIleLeuAspThrValHisSerCysGluA-gAlaA-gGluLeuGlnLysAspHis 925
 QY 789 TGTGAACACGTACAGGAAATGAAACAAATGAAAGAGATTTGTGTACTGAAAG 848
 Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 945
 QY 849 AAACGTGTCAGAGCAAAAGAAATAAATACAGTTAGAGAACCAAAAGTTAAATGGGAA 908
 Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla 985
 QY 969 GATATATTAATCAAAATTTAGGAGAAATAGGAAGATTCAGAGACCATAGGAA 1028
 Db 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAGTCAAAACAACTTCAACAGGCTCTCAGAAATACAGATATAGAAATGGAAG 1088
 Db 1006 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTAGAAAGTAAATTTGATCAGGTTCTCTCAGCTCATGAAATGAAATATATCTCTTA 1148
 Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnThrLeuLeu 1045
 QY 1149 CATGAAATTTGATGTTGAAAGGAAATTTGCGATGCTTAAACTGGAATAGCCACACTG 1208
 Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAACACCAATACCGAGAAAGAAATAAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085

QY 1269 AAGAATGCTGAATCTCAGATGACCCCTAAATCTGAAAGAGAAATCATTAATAAAGGCA 1328
 Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTCTCATGCTGAGAACACAACTGCTCCTCTTAA 1388
 Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGGAAACAAAGACAAAGAAATCTAGAGCGAGAAATTTGAATCACAACCTCTAGA 1448
 Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
 QY 1449 CTGGCTTCTCTCTACAGACCATGATCAAAATGTCACATCAAGAAAGTCAAGACCT 1508
 Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCCACATTCAGAGAGATGCTTTGTCAAAGAAATGATGTTGATGTGAGTAGT 1568
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValSerSer 1185
 QY 1569 ACCGATATATAAATGAGGTGCTCATCAACCACTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSer 1205
 QY 1629 AAGCTAAATTAATCTCAATTCAGAGATGCTCTAAGAGAAATACATTCGTTTC 1688
 Db 1205 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
 QY 1689 AGGAACATGCAAAAGAGACCAACGTGAAACACAGTGTCAATGAAAGGAGCTGTAACAC 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 1245
 QY 1749 TGINTCAAAACGACAGATGATGTGAACAAACACACACTGANCAGAGGAGTCTCTGATC 1808
 Db 1245 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAATTTATTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCCAC 1868
 Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
 QY 1869 ATAANGAAGCTGACACAAACCAAGATTAACAAATGATTTTTCATTTCTTGAGAGAA 1928
 Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
 QY 1929 ATGCN-CATCATCTCTTAAAGAGAAATGAGAGATATTNNATTACNATAACCATTTA 1987
 Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
 QY 1988 AAAACCCGTATATTCATATTCGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2027
 Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337
 RESULT 4
 ABR47548
 ID ABR47548 standard; protein; 1341 AA.
 XX AC ABR47548;
 XX DT 12-JUN-2003 (first entry)
 XX DE Breast cancer associated protein sequence SEQ ID NO:334.
 XX KW Human; breast cancer; cytostatic; gene therapy.
 XX OS Homo sapiens.
 XX PN WO2003004989-A2.
 XX PD 16-JAN-2003.
 XX PF 21-JUN-2002; 2002WO-US019669.
 XX PR 21-JUN-2001; 2001US-0299887P.

PR 27-JUN-2001; 2001US-0301572P.
 PR 18-JUL-2001; 2001US-0306501P.
 PR 25-SEP-2001; 2001US-0325002P.
 PR 05-MAR-2002; 2002US-0362585P.
 PR 14-MAY-2002; 2002US-0380391P.
 XX
 XX (MILL-) MILLENIUM PHARM INC.
 XX
 XX Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;
 PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
 PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
 XX
 XX WPI; 2003-210381/20.
 DR N-PSDB; ACC50246.
 XX
 XX Breast cancer diagnosis or treatment by comparing the level of expression
 PT of a marker in a patient sample with that in the control non-breast
 PT cancer sample.
 XX
 PS Claim 1; SEQ ID NO 334; 128pp; English.
 XX
 XX The present invention describes a method for assessing whether a patient
 CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of
 CC expression of the marker in a control non-breast cancer sample, where a
 CC significant increase in the level of expression of the marker in the
 CC patient sample and the normal level is an indication that the patient is
 CC afflicted with breast cancer. The breast cancer associated sequences from
 CC the present invention have cytostatic activities and can be used in gene
 CC therapy. The method is useful for diagnosing and treating breast cancer.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1341 AA;
 Alignment Scores:
 Pred. No.: 2,04e-274 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 6 Gaps: 0
 US-09-602-362E-15 (1-2030) X ABR47548 (1-1341)
 QY 9 GTTAAAGATGGTCTTCTGAGGCTAACTCGGAATGAAGATTCTTATTCACCTAAAGCC 68
 DB 666 VallyAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
 QY 69 TTAGATTGATGACATGCAACTTTCAGAGAGAGCCCTCCGAGAGGCACTCTCCCTTC 128
 DB 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705
 QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTCCAAATAAAGCCCTTGAATTGAAGATGAA 188
 DB 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
 QY 189 CAACACTTGAGCAGATGATCTCCCATCAGATCCAAACAAAGGACTATCAAGAA 248
 DB 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspIyrGluGlu 745
 QY 249 AGTTCTTGGGATCTCAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTACCC 308
 DB 746 SerSerTipAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGCTACACATCAAAAGAAATAGATAAATAATGAAATTAAGAGTCTCTGAT 368
 DB 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
 QY 369 AATGATGGTCTTCTGAGGCTCCCTGCAGATCAAAAGTTCTTATTCCAACTAAAGCCTTA 428

DB 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAGGCATCTGCTTCGAG 488
 DB 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CTGCTCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATGAAGATGAACAA 548
 DB 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACACATGAGTTCCTTCAGATCAAAACAAAGAGAGGTGTAAGAAAAT 608
 DB 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGATTCTCAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCAAG 868
 DB 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGGAATAAAGTGAAGAAATAGAGATTCACATCACTACCTA 728
 DB 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluLysAspSerThrSerLeu 905
 QY 729 TCAAAAATCTGGATACAGTTCATTCTTGTGAAGAGCAAGGGAACCTTCAAAAAGATCAC 788
 DB 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
 QY 789 TGTGAACAACGTACAGAAAATGGAACAATAAGAAAGAGTGTGTGTACTGAAAAG 848
 DB 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 945
 QY 849 AAATCTGTCAAGCAAAAGAAATAAATCACAGTTAGAGAACCAAAAGTTAAATGGGAA 908
 DB 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTCTGAGTGTGAGATTGATTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 DB 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
 QY 969 GATATATTAATCAAAAATTTAGGGAAGATTAAGGAAGATCAAGAGAGCATAGAAA 1028
 DB 986 AspIleLeuAsnGluLysIleArgGluLeuLeuGlyArgIleGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAACTGAACAACAACTTGAACAGGCTCTCAGAAATACAGATATAGAAATGAAG 1088
 DB 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTCTAGAAATTAATTAATGATCAGGTTCTTCACACTCATCAAAATGAAATATCTCTTA 1148
 DB 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnIyrLeuLeu 1045
 QY 1149 CATGAAATTCATGTTGAAAAAGAAATTCGCATCTCAAACTGAAATAGCACACATG 1208
 DB 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAACACCAATACAGGAAGAAATTAATTAATCTTTGAGCAGATTAAGATTTTAAAGAA 1268
 DB 1066 LysHisGlnIyrGlnGluLysGluAsnLysIyrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAATCTCTCAACTCTCAGATGACCTTAAACTCAAAAGAGAGAAATCAATAAAGGGCA 1328
 DB 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysAsnGala 1105
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTCTTGATAGCTGAGACACATGCTCACTTCTAA 1388
 DB 1106 SerGlnIyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGGAAAAACAAGCAAAAGAAATAGAGCCAGAAATTAATCAACACCATCTCTAGA 1448
 DB 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
 QY 1449 CTGCTTCTCTGTACAGACCATGATCAAAATTTGACATCAAGAAAGTCAAGAACCT 1508
 DB 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165

```

QY 1509 GCTTCCACATTCGAGGAGCTGTTGTTGCAAGAGAAAAATGAATGTTGATGTGAGTAGT 1568
Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATACATGAGTGTCTCCATCAACCACTTTCTGAAGCTCAAGAGAAATCCAN 1628
Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 1205
QY 1629 AAGCCTAAATAATCTCAATATGTCAGGAGATGCTCTAAGAGAAAAATACATTTGTTTC 1688
Db 1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
QY 1689 AGGACATCCACAAAGAGACCACGTCGAACACAGCTGTCGAATGAGGAGCTGAACACA 1748
Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
QY 1749 TGTNTCAAAACCAAGATNATGCAACAAACACACTCANAGCAGGAGTCTCTAGATC 1808
Db 1245 eTyrGluAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY 1809 AGAATATTTCACTACAAAGCAAAATATATGTGGCTTCAACAGCAATGATTCATGCAC 1868
Db 1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATAACAATTTGATNTTTCATTNTCTTGAGAGAAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTTCTTAAAGAGAAAAATAGGAGATATTNATTACNATAACCATTTA 1987
Db 1305 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAAAACCGTATATTCATATGGAAGAAAAAANAAAAA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyr-GluLysGluLysAlaGlu 1337

RESULT 5
ID ABJ37788
AB 1349 AA.
XX AC ABJ37788;
XX AC ABJ37788;
DT 15-MAY-2003 (first entry)
XX Human tumour-related protein - SEQ ID No 573.
DE Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
KW Tumour; breast cancer; cancer; immune response stimulation.
XX OS Homo sapiens.
XX WO200283956-A1.
XX 24-OCT-2002.
XX 15-APR-2002; 2002WO-US012378.
XX 13-APR-2001; 2001US-00834759.
XX 07-DEC-2001; 2001US-00007805.
XX 13-FEB-2002; 2002US-00076622.
XX (CORI-) CORIXA CORP.
XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
PI Vedvick TS, McNeill PD, Durham M;
XX WPI; 2003-103376/09.
XX New polypeptide and polynucleotide useful for stimulating and/or
PT expanding T cells specific for a tumor protein and treating breast
PT cancer.

```

```

XX Example 12; Page 353-357; 375pp; English.
XX The invention comprises a method of stimulating and/or expanding T cells
CC specific for a tumour protein. The invention further comprises human
CC nucleic acids and proteins that are associated with tumours (e.g. breast
CC cancer). The method and sequences of the invention are useful for
CC stimulating and/or expanding T cells specific for a tumour protein,
CC detecting the presence of cancer, stimulating an immune response in a
CC patient and treating breast cancer. The present amino acid sequence
CC represents a human tumour-related protein.
XX Sequence 1349 AA;

Alignment Scores:
Pred. No.: 2,04e-274 Length: 1349
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: Gaps: 0

US-09-602-362E-15 (1-2030) x ABJ37788 (1-1349)
QY 9 GTTAAAGATGCTCTCTGAAGGCTAACTGCGGAATGAAAGCTTCTATTCCAACTAAAGCC 68
Db 674 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 693
QY 69 TTGAATGTATGACATGCAAACTTTTAAAGCAGAGCCTCCCGAGAGCCATCTGCTTC 128
Db 694 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 713
QY 129 GAGCCTCCCACTTCAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAAATGAAGATGA 188
Db 714 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 733
QY 189 CAAACATGTGAGCAGATGAGATCTCCCATCAGATCCCAACAAAGGACTATGAAGAA 248
Db 734 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 753
QY 249 AGTCTCTGGATTCTGAGAGTCTCTGAGACTGTTTCAGAGAGGAGTGCTGTTACCC 308
Db 754 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 773
QY 309 AAGCTACACATCAAAAAGAAATAGATATAAATAAATGGAATAATAGAAAGTCTCTGAT 368
Db 774 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 793
QY 369 AATGATGTTTCTGAAGGCTCCCTCGAGATGAAAGTTTCTATTCCAACTAAAGCTTGA 428
Db 794 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 813
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAGCCATCTGCTTCGAG 488
Db 814 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 833
QY 489 CCTGCCATTCAAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGAATTTGAAGATGAACAA 548
Db 834 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 853
QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAGAGAAAT 608
Db 854 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 873
QY 609 TCTTGGATTCTGAGAGTCTCCCTGAGACTGTTTCACAGAGGATGTGTGTACCCAAAG 668
Db 874 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 893
QY 669 CTTACACATCAAAAAGAAATGGAATAAATAAGTGGAAATTAGAAGATTCACACTAGCCTA 728
Db 894 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 913
QY 729 TCAAAAATCTTGGATACAGTTCAATCTTGTGAAAGAGCAAGGGAAGTTCAAAAAAGATCAC 788

```

Db 914 SerLysIleLeuAspThrValHisSerCysGluuArgAlaArgGluLeuGlnLysAspHis 933
 QY 789 TGTGAACAACTGACAGGAAATGGAACAAATGAAAGAAAGATTTTGTGTACTGAAAG 848
 Db 934 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 953
 QY 849 AACTGTCTAGAGCAAAAGAAATATAATCACAGTTAGAGAACCAAAAGATTAAATGGGAA 908
 Db 954 LysLeuSerGluLysLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 973
 QY 909 CAAGAGCTCTGAGTGTGAGTGAATTTAAACCAAGAAAGAGAGAGAGAAATGCC 968
 Db 974 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluArgGlnAla 993
 QY 969 GATATATTAAATGAAATATAGGAGAAATAGGAGAAATAGGAGAGAGAGAGAGAGAG 1028
 Db 994 AspIleLeuAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1013
 QY 1029 GAGTTAGAGTGAACCAACACTTGACAGGCTCTCAGAAATACAGATATAGAAATGAAG 1088
 Db 1014 GluLeuGluValLysGlnGlnLeuGlnAlaLeuArgIleGlnAspIleGluLys 1033
 QY 1089 AGTGTAGAAGTAATTTGATCAGCTTTCTCAGCTCATGAAATCAAAATATCTCTTA 1148
 Db 1034 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1053
 QY 1149 CATGAAATTTGCAATGTTGAAAAAGGAAATGTCATGCTAAAACTGAAATAGCCACCTG 1208
 Db 1054 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluLeuAlaThrLeu 1073
 QY 1209 AAACACAAATACAGGAAAGGAAATATAATCTTTGAGGACATTAAGATTTAAAGAA 1268
 Db 1074 LysHisGlnTyrGlnGlnLysGluAsnLysTyrPheGluAspIleLysLysLeuLysGlu 1093
 QY 1269 AGAAATGCTGAATCTCAGATGACCTTAAATCTGAAAGAGAAATCAATTAATAAGGGCA 1328
 Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLysLysLysLysLysLysLysLysLys 1113
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGTGAACACAAATGCTCTCTTAA 1388
 Db 1114 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1133
 QY 1389 TTGAAGCAAAACACAGCAAGAAATCTAGAGGAGAAATGAATCAACACCATCTTGA 1448
 Db 1134 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluLeuGluSerHisProArg 1153
 QY 1449 CTGGCTTCTGCTACAGACCATGATCAATTTGATGACATCAAGAAAGTCAAGAACT 1508
 Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
 QY 1509 GCTTTCACATTCAGAGAGATGCTTTCGCAAGGAAATGAATGCTGATGAGTAGT 1568
 Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1193
 QY 1569 ACCGATATATAAATCAGAGTCTCCATCAACACCTTCTGAGCTCAAGAGAAATCCAN 1628
 Db 1194 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
 QY 1629 AAGCTTAAATTAATCTCAATATGACAGAGATGCTTCAAGAGAAATACATTTGTTTC 1688
 Db 1213 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1233
 QY 1689 AGCAACATGACAAAGAGACCAACAGTGAACACAGTGTCAATGAAGAGAGAGTGAACACA 1748
 Db 1233 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
 QY 1749 TGTNTCAAACCAAGATATGATGAAACAAACACTGANCAGCAGGAGTCTTAGATC 1808
 Db 1253 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 1273
 QY 1809 AGAAATTTTCAACTACAAGCAAAATATGTCGCTTCAACAGCAATTAGTTTCATGCAC 1868

Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaH 1293
 QY 1869 ATRANGAAAGCTCACAAACAAAGCAAGATAAATGATTTTCATTTCTTGAGAGGAAA 1928
 Db 1293 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1312
 QY 1929 ATGCN-CATCATCTTCTAAAGAGAAATAGAGAGATTTTATTTACATATACCATTTA 1987
 Db 1313 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1332
 QY 1988 AAAAAACCCGATATTTCAATATGCAAAAAAANAANAANA 2027
 Db 1333 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1345
 RESULT 6
 AAU33351
 ID AAU33351 standard; protein; 1002 AA.
 XX AC AAU33351;
 XX 18-DEC-2001 (first entry)
 XX Human breast cancer protein B726P from alternatively spliced cDNA.
 XX Human; ss; breast cancer protein; tumour; cancer; cytostatic;
 XX Gene therapy; immunogen.
 XX Homo sapiens.
 XX WO200179286-A2.
 XX 25-OCT-2001.
 XX 12-APR-2001; 2001WO-US012164.
 XX 17-APR-2000; 2000US-00551621.
 XX 08-JUN-2000; 2000US-00590751.
 XX 22-JUN-2000; 2000US-00604287.
 XX 20-JUL-2000; 2000US-00620405.
 XX (CORI-) CORIXA CORP.
 XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
 WPI; 2001-611721/70.
 XX N-PSDB; AAS47411.
 XX Breast Tumor Proteins and nucleic acids useful for the prevention,
 diagnosis and treatment of breast cancer.
 XX Claim 3; Page 281-283; 297pp; English.
 The invention relates to isolated breast tumour proteins and nucleic acids that encode them, including immunogenic fragments of the proteins. Also included are expression vectors expressing the proteins, transformed cells and antibodies raised against the proteins or an antigen presenting cell expressing the protein. The proteins and nucleic acids may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate breast tumour protein expression, i.e. breast tumours and breast cancer e.g. by gene therapy. The nucleic acids and their complements may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. The proteins, nucleic acids and antibodies may be used in assays to identify modulators (e.g. antagonists) of breast tumour protein expression and activity. The antibodies and antagonists may also be used to down regulate expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of the proteins in samples (e.g. by enzyme linked immunosorbent assay (ELISA)) and in other immunopurification diagnostic techniques. The present sequence is a breast tumour protein encoded by a cDNA from a breast tumour cDNA library isolated by subtractive hybridisation against a normal breast CDNA library

XX	SQ	Sequence	1002 AA;
Alignment Scores:			
Pred. No.:	1,66e-273	Length:	1002
Score:	3261.00	Matches:	654
Percent Similarity:	97.19%	Conservative:	3
Best Local Similarity:	96.75%	Mismatches:	15
Query Match:	95.83%	Indels:	5
DB:	4	Gaps:	0
US-09-602-362E-15 (1-2030) x AAU33351 (1-1002)			
QY	9	GTATAAGATGCTTCTCGAAGCTTAACCTCCGGAAACAAGTTTCATTCCCACTAAAGCC	68
DB	327	VAllyAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla	346
QY	69	TTAGAATTGATGCACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAAGCCCATCTGCCCTC	128
DB	347	LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe	366
QY	129	GAGCCTGCCATTGAATGCAAAAGCTGTTCCAAATAAGCCCTTGGAATTCGAAGATCAA	188
DB	367	GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu	386
QY	189	CAAAACATTCAGACACATGAGATCTCCCATCAGAACTCAAAACAAAAGAGACTATGAAGAA	248
DB	387	GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu	406
QY	249	AGTCTTGGATTCTCAGAGTCTCTGTGAGATCTGTTTACAGAAGATGTGTGTTACC	308
DB	407	SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro	426
QY	309	AAGCCTACACATCAAAAGAAATAGATAAATAAATGGAATATTAGAAGAGTCTCTGAT	368
DB	427	LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProasp	446
QY	369	AATGATGGTTTTCTGAAGCTCTCCGACAAATGAAAGTTTCTATTCCAACATAAGCCCTTA	428
DB	447	AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu	466
QY	429	GAATTGATGCACATCAAACTTTCAGAGCAGAGCCCTCCGAGAAGCCATCTGCCCTTCGAG	488
DB	467	GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu	486
QY	489	CCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGCAATTCGAATTCGAACAA	548
DB	487	ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln	506
QY	549	ACATTGAGACGACATCAGATGTTCCCTTCAGAAATCAAAACAAAAGAGGTTGAGAAAAAT	608
DB	507	ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn	526
QY	609	TCTTGGATTCTCAGAGTCTCCGTGAGATCTTTCACAGAAGATGTGTGTACCCCAAG	668
DB	527	SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProlys	546
QY	669	GCTACACATCAAAAGAAATGGAATAAATAGCTGGAATAATTAGAAGATTCACACTAGCCTA	728
DB	547	AlaThrHisdGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu	566
QY	729	TCAAAATCTTGATACAGTTCATCTCTGTGAAGCGAAGGAACTTCAAAAGATCAC	788
DB	567	SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis	586
QY	789	TGTGAACAACTACAGCAAAATGGAACAAATGAAAGAAAGCTTTTGTGTACTGAAAG	848
DB	587	CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys	606
QY	849	AACTGTCAGAACCAAAAGAAATAAATCACAGTTAGAACCAAAAGTTAAATCGGAA	908
DB	607	LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu	626
QY	909	CAAGAGCTCTGCAGTCTGAGATTGACTTTTAAACCAAGAGAGAGAGAGAGAAATGCC	968
DB	627	GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluGluLysArgAsnAla	646
QY	969	GATATATTAAATGAAAAAATAGGGAAGAATTAGGAAGATCGAAGACAGCATAGGAAA	1028
DB	647	AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys	666
QY	1029	GAGTTAGAAGTGAACAACAACACTTCGAACAGGCTCTCAGATACAAGATATAGATTGAG	1088
DB	667	GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLys	686
QY	1089	AGTGAGAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAAAATGAAAAATATCTCTTA	1148
DB	687	SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrrLeuLeu	706
QY	1149	CATGAAAAATTGATGTTGAAAAAGAAATTCCTCATGCTTAAACCTGGAAATAGCCACACTG	1208
DB	707	HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu	726
QY	1209	AAACACCAATACCCAGAAAAAGAAAAATAATACTTTTGAAGACATTAAAGATTTTAAAAAGAA	1268
DB	727	LysHisGlnTyrrGlnGluLysGluAsnLysTyrrPheGluAspIleLysIleLeuLysGlu	746
QY	1269	AAAGATTCGTGAACTTTCAGATGACCCCTTAAACCTGAAAGAGGAATCATTAACATAAGGCCA	1328
DB	747	LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla	766
QY	1329	TCTCAATATAGTGGCAGCTTAAAGTTTCTGATAGCTGAGAACACAAGTCTCCTCTCTAAA	1388
DB	767	SerGlnTyrrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys	786
QY	1389	TTGAAGAAAAACAAGACAAAGAAATACTAGAGGCAGAAAAATTGAATCACACCATCTCTAGA	1448
DB	787	LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg	806
QY	1449	CTGGCTTCGTGTTACAAGCACCATGATCAAAATGFGACATCAAGAAAAAGTCAGAACCT	1508
DB	807	LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro	826
QY	1509	GCTTTCCACATTGCAGGAGATGCTTTGCAAGAAAAAATGAATGTTGATGTCAGTAGT	1568
DB	827	AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer	846
QY	1569	ACCGATATATAACAATGAGGTCTCCATCAACCCACTTCTTCAAGAGCTCAAGAGAAATCCAN	1628
DB	847	Thr-IleTyrrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy	866
QY	1629	AAGCTTAAAAATTAATCTCAATTATGCAGGAGATGCTCTAAGAGAAATATACATTGGTTTC	1688
DB	866	sSerLeuLysIleAsnLeuAsnTyrrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe	886
QY	1689	AGGAACATGCACAAAGAGACCAACCTGAAACACACAGTGTCAATGAAGAGAACTGAACACA	1748
DB	886	r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM	906
QY	1749	TGNTCAAANCACACAGATNATGTGAACAAACACACATGANCCAGCAGGAGTCTCTAGATC	1808
DB	906	etTyrrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG</	

QY 1149 CATGAATTTGCGATGTTGAAAGGAAATTCCTCATCTGCTAAAGCTGGAATAGCCACACTG 1208
 Db 707 HisGluAsnCysMetLeuLysGluLeuAlaMetLeuLysLeuGluLeuAlaThrLeu 726
 QY 1209 AAACACCAATACACAGGAAAGGAAATTAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AGAATGCTGAATCTCAGATGACCTAAACCTGAAAGAGGAATCATTAACATAAGGGCA 1328
 Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTCTGTATAGCTGAGACACAAATCTCTCACTCTAAA 1388
 Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuLeuAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAAGGAAACAAAGCAAGAAATACTAGAGGAGAAATTAATCAACCATCTCTAGA 1448
 Db 787 LeuLysGluLysGlnAspLysGluLeuLysGluAlaGluLeuGluSerHisHisProArg 806
 QY 1449 CTGGCTTCTGTGTACAGACCATGATCAATTTGACATCAAGAAAGTCAAGACCT 1508
 Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTTCCACATTCAGAGAGATGCTTTGTTGCAAGAAAAATCAATGTTGATGTGAGTAGT 1568
 Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATACATGAGTGCTCCATCAACCATCTTCTGAAGCTCAAGGAATCCAN 1628
 Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 866
 QY 1629 AAGCCTAAAAATTAATCTCAATATTGTCAGGAGATGCTCTAAGAGAAAAATACATTGGTTTC 1688
 Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
 QY 1689 AGGAACATGCAAAAGAGACCAACCTGAAACACAGTGTCAAATGAAGAACTGAACACA 1748
 Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAANGCAACAGATNATGGAACACACACACACACACAGGAGCTCTAGATC 1808
 Db 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
 QY 1809 AGAAATTTTCACTACAAAGCAAAATATGTGCTTTCAACAGCAATAGTTTCATGCAC 1868
 Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaH 946
 QY 1869 ATAANGAAGCTGACAAAGAGCAGATACAAATGATNTTCATTNTCTTTCAGAGGAA 1928
 Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCN-CATCATCTTTCTAAAAGAGAAAAATGAGGAGATATTNTTATCATNPAACCATTTA 1987
 Db 966 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAACCCGTATATTTCATATGGAAGAAAAAANAAAAA 2027
 Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 8

ID ABU37741
 XX ABU37741 standard; protein; 1002 AA.
 AC
 XX ABU37741;
 DT 15-MAY-2003 (first entry)
 DE Human tumour-related protein - SEQ ID NO 475.
 XX Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
 KW tumour; breast cancer; cancer; immune response stimulation.

XX Homo sapiens.
 OS
 XX WC200283956-A1.
 PN
 XX 24-OCT-2002.
 PD
 XX 15-APR-2002; 2002WO-US012378.
 PP
 XX 13-APR-2001; 2001US-00834759.
 PR 07-DEC-2001; 2001US-00007805.
 PR 13-FEB-2002; 2002US-00076522.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
 PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
 PI Vedvick TS, McNeill PD, Durham M;
 XX WPI; 2003-103376/09.
 DR
 XX New polypeptide and polynucleotide useful for stimulating and/or
 PT expanding T cells specific for a tumor protein and treating breast
 PT cancer.
 XX
 XX Example 1; Page 298-300; 375pp; English.
 PS
 XX The invention comprises a method of stimulating and/or expanding T cells
 CC specific for a tumor protein. The invention further comprises human
 CC nucleic acids and proteins that are associated with tumors (e.g. breast
 CC cancer). The method and sequences of the invention are useful for
 CC stimulating and/or expanding T cells specific for a tumor protein,
 CC stimulating the presence of cancer, stimulating an immune response in a
 CC patient and treating breast cancer. The present amino acid sequence
 CC represents a human tumour-related protein.
 XX
 XX Sequence 1002 AA;

Alignment Scores:
 Pred. No.: 1,66e-273 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 6 Gaps: 0

US-09-602-362E-15 (1-2030) x ABU37741 (1-1002)
 QY 9 GTTAAAGATGCTCTCTGAAGGCTAACTGCGAATGAAAGTTTCTATTCCAACTAAAGCC 68
 Db 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAGCCATCTGCCTTC 128
 Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 366
 QY 129 GAGCCTGCCATTCGAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAATGAGAAATGAA 188
 Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAAACATTTGAGACAGATGAGATCTCCCATCAGAAATCCCAACAAAGACATATGAAGAA 248
 Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTCTTGGGATCTGAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTTACCC 308
 Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAAATGGAATAATAGAGAGTCTCTCGAT 368
 Db 427 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
 QY 369 AATGATGGTTTCTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428

Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTGATCGACATGCAAACTTTCAAGCAGAGCCCTCCGAGAACCCATCTGCCTTCGAG 488
Db 467 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 486
QY 489 CCTGCCATTGAATGCAAAAGTCTGTTCCAAATAAGCCCTTGAATTGAAGNATGAACAA 548
Db 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAAAGTTGAGAAAT 608
Db 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATTCGAGAGTCTCGTGAGACTGTTTCACAGAAGGATGTGTGTATCCCAAG 668
Db 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACACATCAAAAGAAATGGAATAAAGTGAATAATTAGAAGATTCAACTAGCCCTA 728
Db 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
QY 729 TCAAAATCTTGATACAGTTCATCTTGTGAAAGCAAGGGAACCTTCAAAAGATCAC 788
Db 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGAACAACGTACAGGAAATGGAACAAATGAAAGAAAGTGTGTACTGAGAAAG 848
Db 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 606
QY 849 AAACCTGTCAAGCAAAAGAAATAAATCACAGTTAGAGAACCAAAAGTTAAATGGAA 908
Db 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
QY 909 CAAGACTCTGAGTGTGAGTGTACTTAAACCAAGAGAGAGAGAGAGAGATGCC 968
Db 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
QY 969 GATATATAAATCAAAATAGGAAGAAATAGGAAGAAATAGGAAGCAAGCATAGGAAA 1028
Db 647 AspIleLeuAsnGluLysIleArgGluLeuGluArgIleGluGluGlnHisArgLys 666
QY 1029 GAGTTAGAGTGAACAACAACCTGAAAGGCTCTCAGATACAGATACAGATAGAG 1088
Db 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLys 686
QY 1089 AGTGTAAGAAATTAATTTGAATCAGGTTCTCACACTCATGAAATGAAATATCTCTTA 1148
Db 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
QY 1149 CATGAAATTCATGTTGAAAAGGAATGTCATGCTTAAACTGGAATAGCCACATG 1208
Db 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
QY 1209 AAACACCAATACCAAGAAAGGAATAATACTTTGAGGACATTAAGATTTTAAAGAA 1268
Db 727 LysHisGlnTyrGlnGlnLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
QY 1269 AAGAATCTGAACTTCAGATGACCCCTAAACCTGAAAGAGGAATCAATTAACATAAGGCA 1328
Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLysSerLeuThrLysArgAla 766
QY 1329 TCTCAATATAGGGCAGCTTAAAGTCTGATAGCTGAGAACACATGCTCACTCTTAA 1388
Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGAAAAACAACACAAAGAAATCTAGAGGAGAAATTAATCAACATCCTAGA 1448
Db 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 806
QY 1449 CTGGCTCTCTGTACAAAGACCATGATCAATGTGCATCAAGAAAGTCAGAACT 1508

Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCCACATTCGAGAGATGCTTGTTCGCAAGAAAAATGAATCTTATGATGAGTAGT 1568
Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACAAATGAGGTCTCCATCAACACATTTCTGAAGCTCAAGGAAATCCAN 1628
Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
QY 1629 AAGCTTAAATTAATCTCAATTCAGAGAGATGCTTAAGAGAAAAATACATTGCTTTC 1688
Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
QY 1689 AGGAACATGCACAAAGAGACCAACGCGAAACACAGTGTCAAATGAAGGAACTGAACACA 1748
Db 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
QY 1749 TGTNTCAANCGAACAAAGATNATGTCAACAAACACTCANCAAGCAGAGTCTCTAGATC 1808
Db 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
QY 1809 AGAAATTTTCACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTCATGCAC 1868
Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
QY 1869 ATAAGCAAGCTGCACAAACAAAGCAAGATAACAAATGTATNTTCATTNTCTTCAGAGAAA 1928
Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
QY 1929 ATGCN-CATCATCTTCTTAAAGAGAAAAATGAGGACATATTNTTACNATAACCATTTA 1987
Db 966 MetGlnHisLeuLeuLysGluLysAsnGluIlePheAsnTyrAsnAsnHisLeu 985
QY 1988 AAAAACCCGCTATATTCAATATGCAATATGCAAAAAAANAANAANA 2027
Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998
RESULT 9
ADE44427
ID ADE44427 standard; protein; 1002 AA.
AC ADE44427;
XX 29-JAN-2004 (first entry)
DT Human breast cancer protein #10.
DE human; breast tumour; cancer; vaccine; T cell stimulator;
KW T cell expander.
XX Homo sapiens.
OS US2003104366-A1.
FN 05-JUN-2003.
PD 17-APR-2000; 2000US-00551621.
PF 28-DEC-1998; 98US-00222575.
PR 02-APR-1999; 99US-00285480.
PR 23-JUN-1999; 99US-00339338.
PR 02-SEP-1999; 99US-00389681.
PR 03-NOV-1999; 99US-00433826.
XX (JIAN/) JIANG Y.
PA (DILL/) DILLON D C.
PA (MITC/) MITCHAM J L.
PA (XUJJ/) XU J.
PA (HARL/) HARLOCKER S L.
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL;

DR WPI; 2004-020270/02.
 DR N-PSDB; ADE44426.
 XX Novel isolated polypeptide comprising immunogenic portion of breast tumor
 PT protein or its variant, useful for formulating vaccines for inhibiting
 PT cancer development in a patient.
 XX

PS Example 1; SEQ ID NO 475; 217pp; English.

XX The invention relates to an isolated polypeptide comprising at least an
 CC immunogenic portion of a breast tumour protein. The polynucleotide, its
 CC polypeptide, its antibody, a pharmaceutical composition comprising the
 CC fusion protein or the polynucleotide encoding it, a vaccine comprising
 CC the fusion protein or the polynucleotide encoding it, an isolated T cell
 CC population comprising T cells specific for a breast tumour protein, and a
 CC method for removing tumour cells from a biological sample is useful for
 CC inhibiting the development of a cancer in a patient. The polypeptide is
 CC useful for stimulating and/or expanding T cells specific for a breast
 CC tumour protein. Stimulating and/or expanding T cells specific for a
 CC breast tumour protein is useful for inhibiting the development of a
 CC cancer in a patient. The method additionally involves the step of cloning
 CC at least one proliferated cell and then administering the cloned T cells
 CC to the patient. The present sequence represents a breast cancer protein.
 XX

SQ Sequence 1002 AA;

Alignment Scores:
 Pred. No.: 1.66e-273 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 8 Gaps: 0

US-09-602-362E-15 (1-2030) x ADE44427 (1-1002)

QY 9 GTTAAAGATGGCTCTCTGAGGCTAACTCGGAATGAAAGTTCTATTCCAACTAAAGCC 68
 DB 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 59 TTAGAAATGATCGCATCGAACTTCAAGCAGAGCCCTCCGAGAGCCATCTCCCTTC 128
 DB 347 LeuGluLeuMetAspMetGlnThrPhenylsAlaGluProGluGluProSerAlaPhe 366
 QY 129 GAGCTGCTGCAATGAAATGAAAGCTGTCTCCAAATAAAGCCTTGGAAATGAAAGTAA 188
 DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAACATTCAGAGCAGATGAGTACTCCATCAGATTCCAACAAAGAGGACTATCAAGAA 248
 DB 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGGATTCTGAGAGTCTCTGAGACTGTTTCACAGAGGATGTGTGTTTACCC 308
 DB 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATAATGGAATTAAGACAGTCTCCTGAT 368
 DB 427 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
 QY 369 AATGATGTTTCTGAGGCTCCCTCGAGAATGAAAGTTTCTATTCCAACTAAAGCCCTTA 428
 DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
 QY 429 GAATTGATGAGATGCAAACTTTCAAGCAGAGCCCTCCGAGAGCCATCTGCTCCGAG 488
 DB 467 GluLeuMetAspMetGlnThrPhenylsAlaGluProProGluLysProSerAlaPheGlu 486
 QY 489 CCTGCCATTCGAATGCAAAAGCTGTCTCCAAATAAGCCTTGGAAATGAGATCAACAA 548
 DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
 QY 549 ACATTGAGAGCAGATCAGATGTTCCTTCAGAAATCAAAACAAAGAGGTTTCAAGAAAT 608

DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 526
 QY 609 TCTTGGATTCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTATCCCAAG 668
 DB 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
 QY 669 GCTACACATCAAAAGAAATGGATTAATAAGTGGAAATTAGAAGATTCACATAGCCCTA 728
 DB 547 AlathHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
 QY 729 TCAAAATCTTGATACAGTTCATTCTTGTGAAGAGCAAGGAACTTCAAAAAGATCAC 788
 DB 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
 QY 789 TGTGAACAGCTACAGAAAAATGGAACAAATCAAAAGAGCTTTTGTCTACTGAAAGAG 848
 DB 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606
 QY 849 AAATCTGTCAAGCAAAAGAAATATAAATCACAGTTAGAGAACCAAAAGTATAATGGAA 908
 DB 607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
 QY 909 CAAGAGCTGTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 DB 627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
 QY 969 GATATATTAATCAAAATTTAGGAGAGATTTAGGAGAGATTCAGAGAGCAGCATAGAAA 1028
 DB 647 AspIleLeuAsnGlnLysIleArgGluGluLeuGlyArgIleGluGlnHisArgLys 666
 QY 1029 GAGTTAGAACTGAAACCAACTTGAACAGGCTCTCAGATAACAAGATATAGATTGAAG 1088
 DB 667 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
 QY 1089 AGTGTAGAAATTAATTCAGTTCAGTTCCTCACACTCATGAAATGAAATATATCTCTTA 1148
 DB 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnThrLeu 706
 QY 1149 CATCAAAATTCATGTTGAAAAAGGAAATTCCTATGCTAAACTGAAATAGCCACTG 1208
 DB 707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
 QY 1209 AACACCAATACAGGAAAGGAAATAATACTTTGAGCAGATTAAGATTTAAAGAA 1268
 DB 727 LysHisGlnThrGlnGluLysGluAsnLysThrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AAGAACTCTCAACTTCAGATGACCTTAAACTCAAAAGAGAAATCATTAATAAAGGCA 1328
 DB 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATGCTGAGAACACATGCTCCTCTTAAA 1388
 DB 767 SerGlnThrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAGGAAAAACAAGCAAGAAATACTAGAGCAAGAAATTCATCACCACCTCTCTAGA 1448
 DB 787 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 806
 QY 1449 CTGCTCTCTGCTGTACAGACCATGATCAAAATTTGACATCAACAAAACTCAAGAACCT 1508
 DB 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTTCACATTCAGAGAGATGCTTGTTCGCAAGAAATTAAGATGTGTGATGTGAGTAGT 1568
 DB 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAAAGAAATCCAN 1628
 DB 847 Thr-IleThrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
 QY 1629 AAGCTTAAATTAATCTCAATTATGAGAGATGCTCTTGAAGAAAAATACATTGTTTC 1688

Dd 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 886
 QY 1689 AGGAACATGCACAAAGAGACCAACGCTGAAACACAGAGTCAATGAGGAGAGCTGAACACA 1748
 Dd 886 r-GluHisAlaGlnArgGlnArgGlnThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAAAACGACAAAGATNATGTGAACAAACACACACTGACAGCAGGAGTCTCTAGATC 1808
 Dd 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 926
 QY 1809 AGAAATATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATAGTTCATGCMC 1868
 Dd 926 lnyLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnLeuValHisAlaH 946
 QY 1869 ATAANGAAGCTGACACAAACAAAGCAAGATAAACAATGATNTTCATNTCTTGAGAGGAAA 1928
 Dd 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCN-CATCATCTTTCTAAAGAGAAAATGAGAGATATTTNATTACNATTAACCATTTA 1987
 Dd 966 MetGlnHisLysLeuLysGluLysAsnGluLysIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAAAACCGTATATTTCAATATGGAAGAAAAAANAANA 2027
 Dd 986 LysAsn-ArgIleGlnTyrGlnLysGluLysAlaGlu 998

RESULT 10
 AAU33357
 ID AAU33357 standard; protein; 1095 AA.
 AC AAU33357;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human breast cancer protein B726P fusion protein #1.
 XX
 KW Human; ss; breast cancer protein; tumour; cancer; cytostatic;
 KW gene therapy; immunogen.
 XX
 OS Homo sapiens.
 XX
 PN WO200179286-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001WO-US012164.
 XX
 PR 17-APR-2000; 2000US-00551621.
 PR 08-JUN-2000; 2000US-00590751.
 PR 22-JUN-2000; 2000US-00604287.
 PR 20-JUL-2000; 2000US-00620405.
 XX
 PA (CORI-) CORIXA CORP.

PI Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
 XX WPI; 2001-611721/70.
 XX N-PSDB; AAS47421.
 XX
 PT Breast Tumor Proteins and nucleic acids useful for the prevention,
 PT diagnosis and treatment of breast cancer.
 XX
 PS Claim 22; Page 292-295; 297pp; English.

CC The invention relates to isolated breast tumour proteins and nucleic
 CC acids that encode them, including immunogenic fragments of the proteins.
 CC Also included are expression vectors expressing the proteins, transformed
 CC cells and antibodies raised against the proteins or an antigen presenting
 CC cell expressing the protein. The proteins and nucleic acids may be used
 CC in the prevention, diagnosis and treatment of diseases associated with
 CC inappropriate breast tumour protein expression, i.e. breast tumours and
 CC breast cancer e.g. by gene therapy. The nucleic acids and their
 CC complements may also be used as DNA probes in diagnostic assays to detect

CC and quantitate the presence of similar nucleic acids in samples, and
 CC therefore which patients may be in need of restorative therapy. The
 CC proteins, nucleic acids and antibodies may be used in assays to identify
 CC modulators (e.g. antagonists) of breast tumour protein expression and
 CC activity. The antibodies and antagonists may also be used to down
 CC regulate expression and activity. The antibodies may also be used as
 CC diagnostic agents for detecting the presence of the proteins in samples
 CC (e.g. by enzyme linked immunosorbent assay (ELISA)) and in other immuno-
 CC purification diagnostic techniques. The present sequence is a breast
 CC tumour protein encoded by a cDNA from a breast tumour cDNA library
 CC isolated by subtractive hybridisation against a normal breast cDNA
 CC library
 XX
 SQ Sequence 1095 AA;

Alignment Scores:
 Pred. No.: 1,71e-273 Length: 1095
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: Gaps: 0

US-09-602-362E-15 (1-2030) x AAU33357 (1-1095)
 QY 9 GTTAAAGATGCTCTTCTGAAAGCTAACTCGGAGTGAAGTTTCTATTCTCAACTAAGCC 68
 Dd 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
 QY 69 TTGAGATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGGCATCTGCTCTC 128
 Dd 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
 QY 129 GAGCTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATGAAGATGAA 188
 Dd 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
 QY 189 CAACATTTGAGACAGATGATGATCTCCCATCAGATCCAAACAAAGAGCATTGAGAA 248
 Dd 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
 QY 249 AGTTCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTTCACAGAGGATGTGTGTTCACC 308
 Dd 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
 QY 309 AAGCTACACATCAAAAAGAAATAGATAAATAATGGAATTAAGAGCTCTCTGAT 368
 Dd 520 LysAla***HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
 QY 369 AATGATGTTTCTGAGAGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Dd 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGGCATCTGCCTTCGAG 488
 Dd 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
 QY 489 CCTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATTAAGAGTGAACAA 548
 Dd 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
 QY 549 ACATTGAGACAGATCAGATGTTCCCTTCAGAAATCAAAAACAAAGAGGTTGAGAGAAAT 608
 Dd 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys***ValGluGluAsn 619
 QY 609 TCTTGGGATTCTGAGAGTCTCCGTGAGACTGTTTTCACAGAGGATGTGTGTGTACCCAAG 668
 Dd 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
 QY 669 GCTTACACATCAAAAAGAAATGGAATAAATAAGTGGAAATTAAGAGATTCACATGACCTTA 728
 Dd 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659

729 TCAGAAATCTGGATACAGTTCATCTCTGTGAAAGCAAGCGAAGTTCAGAAAGATCAC 788
Db SerLysileLeuAspThrValHisSerCysGluArgAlaArgLeuGlnLysAspHis 679
789 TGTGAACAAACGTACAGCAAAATGGAACAAATGAAAGAAAGTGTGTCTACTGAAAAAG 848
Db CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 699
849 AACTGTCAGAGCAAAAGAAATATAATCACAGTTAGAGAACCAAAAGTTAAATGGGAA 908
Db LysLeuSerGluAlaLysGluileLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
909 CAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
Db GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739
969 GATATATTAATGAAAAATATAGGGAAGATTAGGAAGATCGAAGAGCAGCATGAGAAA 1028
Db AspIleLeuAsnGluLysileArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 759
1029 GAGTTAGAGTGAACCAACACTTGAACAGGCTCTCAGATACAGATACAGATATGAAATGAG 1088
Db GluLeuGluValLysGlnGlnLeuGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
1089 AGTGTAGAAATGAATTTGAATCAGGTTCTCAGCTCATGAAATGAAATATCTCTTA 1148
Db SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
1149 CATGAAATTCATGTTGAAAGAGAAATGCGATCTGCTAAACTGGAATAGCCACATG 1208
Db HisGluAsnCysMetLeuLysLysGluileAlaMetLeuLysLeuGluileAlaThrLeu 819
1209 AAACACCAATACAGGAAAGAAATAATAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
Db LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysileLeuLysGlu 839
1269 AGAATGCTGAATCTCAGATGACCCCTAAACTGAAAGAGGAATCATTAATAAGGGCA 1328
Db LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 859
1329 TCTCAATATAGTGGGAGCTTAAAGTCTGATAGCTGAGACACAACTGCTCACTTCTAAA 1388
Db SerGlnTyrSerGlyGlnLeuLysValLeuileAlaGluAsnThrMetLeuThrSerLys 879
1389 TTGAGGAAACCAAGACAAAGAAATATAGAGGAGAGAAATTCATATCACCATCCTAGA 1448
Db LeuLysGluLysGlnAspLysGluileLeuGluAlaGluileGluSerHisHisProArg 899
1449 CTGGCTTCTGCTACAGACCATCATCAATGTGACATCAAGAAAGTCAAGAACT 1508
Db LeuAlaSerAlaValGlnAspHisAspGlnileValThrSerArgLysSerGlnGluPro 919
1509 GCTTTCCCATTCAGAGATGCTTTGTCAGAAAGAAATGAATGTTGATGTGAGTAGT 1568
Db AlaPheHisileAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
1569 ACGGATATATACATAGAGTGTCTCATCAACACTTCTGAGCTCAAGAAATCCAN 1628
Db Thr-ileTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerly 959
1629 AAGCCTAAAAATTAATCTCAATATATCAGAGATGCTCTAAGAGAAATACATGTTTC 1688
Db sSerLeuLysileAsnLeuAsnTyrAlaGlyaspAlaLeuArgGluAsnThrLeuValSe 979
1689 AGGAACATGCACAAAGACCAACGTGAACACAGTGTCAATGAAGGAAGTGAACACA 1748
Db r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 999
1749 TGTTNTCAACAGCAAGATNATGTCAACAAACACACTGANCAGCAGGAGTCTTAGATC 1808
Db etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1019
1809 AGAATATTATTTCAACTACAAAGCAAAATATATGGCTTCAACAGCAATTAGTTCATGCAC 1868

Db 1019 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnLeuValHisAlaH 1039
Qy 1869 ATAANGAAAGCTGCAACAAAGCAAGATAACAAATTTCATNTTCTTGAGAGAAA 1928
Db is-LysLysAlaAspAsnLysSerLysileThrileAspIleHisPheLeuGluArgLys 1058
Qy 1929 ATGCN-CATCATCTCTTAAAGAGAGAAAATGAGGAGATATTNATTACNATAACCATTA 1987
Db MetGlnHisHisLeuLeuLysGluLysAsnGluilePheAsnTyrAsnAsnHisLeu 1078
Qy 1988 AAAACCCCGTATATTCAATATGCAAAAAAANAANA 2027
Db LysAsn-ArgileTyrGlnTyrGluLysGluLysAlaGlu 1091
RESULT 11
ABG78924
ID ABG78924 standard; protein; 1095 AA.
XX AC ABG78924;
XX 15-NOV-2002 (first entry)
DT Human breast tumour polypeptide #15.
XX Human breast tumour polypeptide #15.
XX Human; breast tumour protein; breast cancer; cytostatic; vaccine.
XX Homo sapiens.
XX US2002085998-A1.
XX 04-JUL-2002.
XX 13-APR-2001; 2001US-00834759.
XX 28-DEC-1998; 98US-00222575.
XX 02-APR-1999; 99US-00285480.
XX 23-JUN-1999; 99US-00339338.
XX 02-SEP-1999; 99US-00389681.
XX 03-NOV-1999; 99US-00433826.
XX 17-APR-2000; 2000US-00551621.
XX 08-JUN-2000; 2000US-00590751.
XX 22-JUN-2000; 2000US-00604287.
XX 20-JUL-2000; 2000US-00620405.
XX (CORI-) CORIXA CORP.
XX Jiang Y, Dillon DC, Mitcham JL, Xu J, Harlocker SL, Hepler WT;
PI Henderson RA;
XX WPI; 2002-635657/68.
XX N-ESDB; ABS64022.
XX Novel breast cancer polynucleotides and polypeptides encoded by the
PT polynucleotides, useful for detecting the presence of breast cancer in a
PT patient, and in pharmaceutical compositions, for treating breast cancer.
XX Disclosure; Page 223-225; 247pp; English.
XX The invention relates to an isolated breast tumour polynucleotide and the
CC polypeptide it encodes. The polynucleotide and polypeptide are useful for
CC detecting the presence of breast cancer in a patient, and in
CC pharmaceutical compositions for treating breast cancer. The sequences are
CC useful for stimulating an immune response in a patient and can therefore
CC be used in production of vaccines. The sequences are also useful for
CC detecting the presence of a cancer in a patient, by obtaining a
CC biological sample from the patient, contacting the biological sample with
CC a composition of the invention and detecting the amount of polynucleotide
CC that hybridizes to the sample. This sequence represents a human breast
CC tumour polypeptide of the invention
XX Sequence 1095 AA;
SQ

Alignment Scores:

Pred. No.: 1,71e-273 Length: 1095
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservative: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 5 Gaps: 0

US-09-602-362E-15 (1-2030) x ABG78924 (1-1095)

Qy 9 GTTAAAGATGGTCTTCTGAAGGCTAACTGCGGAATGAAGTTTCTATTTCACATTAAGCC 68
 Db 420 ValIysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
 Qy 69 TTAGAATTGATGACATGCAAACTTTCAAGACAGAGCTCCCGAGAGCCATCTGCTTC 128
 Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
 Qy 129 GAGCTGCCATTGAAATGCAAAAGTCTGTTCCTCAAAATAAAGCCCTGGAATGAAGATGAA 188
 Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGlu 479
 Qy 189 CAACACATGACAGATGAGTACTCCCATCAGATCCAAACAAAGGAGCTATGAAGA 248
 Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
 Qy 249 AGTTCTTGGGATTCGAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTTPACCC 308
 Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
 Qy 309 AAGCTACATCAAAAGAAATAGATAAATAAATGGAATTAAGAGTCTCTGTAT 368
 Db 520 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
 Qy 369 AATGATGGTTTCTGAAGGCTCCTCGAGAAATGAAAGTTTCTATTCCAACTAAAGCCCTTA 428
 Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
 Qy 429 GAATTGATGACATGCAAACTTTCAAGACAGAGCTCCCGAGAGCCATCTGCTTCGAG 488
 Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
 Qy 489 CCTGCCATTCAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATGAAGATGAACAA 548
 Db 580 ProIleIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGluGln 599
 Qy 549 ACATTGAGACATGATGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAAGAAAT 608
 Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 619
 Qy 609 TCTTGGATTCGAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTACCCCAAG 668
 Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
 Qy 669 GCTACACATCAAAAGAAATGGATAAATAAGTGGAAATTAAGACATCAACTAGCCTA 728
 Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
 Qy 729 TCANAATCTTGGATACAGTCTTCTTGTGAAAGACGAGGAACTTCAAAAGATCAC 788
 Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGluLysAspPhe 679
 Qy 789 TGTCAACACCTCAGCAAAATGCAAAATGAAAGAAAGTTTGTGTACTGAAAGAAAG 848
 Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
 Qy 849 AAATGTCAGACAGCAAAAGAAATTAATCAAGTTAGAGACCAAAAGTTAAATGGAA 908
 Db 700 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
 Qy 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 720 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739

Qy 969 CATATATTAAATGAAAAAATTAGGAAAGAAATTAGGAAGAAATCGAAGACGACATAGGAAA 1028
 Db 740 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 759
 Qy 1029 GAGTTAGAGTGAACCAACAACTTGAACAGGCTCTCAGATACAGATATAGATATGAAG 1088
 Db 760 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
 Qy 1089 AGTGTAGAAATTAATTTGAATTCAGTTTCTCACACTCATGAAATGAAATTTATCTCTTA 1148
 Db 780 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
 Qy 1149 CATGAAATTTGTCATGTTGAAAGAAATTCGCCATGCTTAAACTGGAATPAGCCACATG 1208
 Db 800 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 819
 Qy 1209 AAACACCAATACCAAGGAAAGGAAATTAATACTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 820 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 839
 Qy 1269 AAGATGCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAATCATTAACATAAAGGCA 1328
 Db 840 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 859
 Qy 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGACACACATGCTCATTCTAAA 1388
 Db 860 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 879
 Qy 1389 TTGAAGGAAAAACAAGCAAGAAATACTAGAGGCAAGAAATTTGAATCACACCATCTCTAGA 1448
 Db 880 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluLeuGluSerHisHisProArg 899
 Qy 1449 CTGCTTCTGCTGACAAAGCCATGATCAAAATGTGTGACATCAAGAAAGTCAAGAACCT 1508
 Db 900 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 919
 Qy 1509 GCTTTCCACATTCGAGGAGATGCTGTTTGCAGAAAAAATGAATGTGTGATGTGAGTAGT 1568
 Db 920 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
 Qy 1569 ACCGATATTAACAAATGAGTGTCCATCAACCACTTCTCAAGCTCAAGGAAATCCAN 1628
 Db 940 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 959
 Qy 1629 AAGCTAAAAATTAATCTCAATATGAGGAGATGCTCTAAGAGAAATACATTGGTTTC 1688
 Db 959 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 979
 Qy 1689 AGGAACATGCAAAAGAGACCAACGTCGAAACACACAGTCTCAAAATGAAGAGAGCTGAAC 1748
 Db 979 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 999
 Qy 1749 TGTNTCAANGCAACAGATNATGTAACAACACACTGANCAGCAGAGTCTCTAGATC 1808
 Db 999 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1019
 Qy 1809 AGAAATTTTCAACTACAAAGCAAAAAATATGTGGCTTCAACAGCAATTAAGTTTCATGCAC 1868
 Db 1019 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnValHisAlaH 1039
 Qy 1869 ATANGAAGCTGACCAACAAAGCAAGCAATTAACATTTGATTTTCTTCTTGAGAGAGAA 1928
 Db 1039 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheAsnHisLeu 1058
 Qy 1929 ATGCN-CATCATCTTCAAAAGAAAAATGAGAGATATTTTATTATTACATAACATTATA 1987
 Db 1059 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1078
 Qy 1988 AAAAACCCGTATTTTCAATATGGAAGAAAAAANAANAANA 2027
 Db 1079 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1091

RESULT 12
ABJ37747
ID ABJ37747 standard; protein; 1095 AA.
XX
XX AC ABJ37747;
XX
XX DT 15-MAY-2003 (first entry)
XX
XX DE Human tumour-related protein - SEQ ID No 493.
XX
XX KW Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
XX tumour; breast cancer; cancer; immune response stimulation.
XX
XX OS Homo sapiens.
XX
XX PN WO200283956-A1.
XX
XX PD 24-OCT-2002.
XX
XX PF 15-APR-2002; 2002WO-US012378.
XX
XX PR 13-APR-2001; 2001US-00834759.
XX
XX PR 07-DEC-2001; 2001US-00007805.
XX
XX PR 13-FEB-2002; 2002US-00076622.
XX
XX PA (CORI-) CORIXA CORP.
XX
XX PI Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
PI Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
PI Vedwick TS, McNeill PD, Durham M;
XX
XX DR WPI; 2003-103376/09.
XX
XX PT New polypeptide and polynucleotide useful for stimulating and/or
XX expanding T cells specific for a tumor protein and treating breast
XX cancer.
XX
XX PS Disclosure; Page 310-312; 375pp; English.
XX
XX CC The invention comprises a method of stimulating and/or expanding T cells
XX specific for a tumour protein. The invention further comprises human
XX nucleic acids and proteins that are associated with tumours (e.g. breast
XX cancer). The method and sequences of the invention are useful for
XX stimulating and/or expanding T cells specific for a tumour protein.
XX detecting the presence of cancer, stimulating an immune response in a
XX patient and treating breast cancer. The present amino acid sequence
XX represents a human tumour-related protein
XX
XX SQ Sequence 1095 AA;

Alignment Scores:
Pred. No.: 1,71e-273 Length: 1095
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 6 Gaps: 0

US-09-602-362E-15 (1-2030) x ABJ37747 (1-1095)

QY 9 GTTAAAGATGGTCTTCTGAGGCTAACTCGGAATGAAAGTTCTATTCCAACTAAAGCC 68
Db 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
QY 69 TTGAATTGATGACATGCAACTTTCAAGCAGAGCCCTCCGAGAGCCATCTGCCTTC 128
Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGGAAATGAAGATGAA 188
Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
QY 189 CAAACATTGAGCAGATGATGATCTCCCATCGAATTCCAACAAAGGACTATGAGAA 248

Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
QY 249 AGTCTCTGGATTCTCAGAGTCTCTGAGACTGTTTTCACAGAAGCATGTGTGTTTACCC 308
Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATAAATGAAAAATTAGAGAGTCTCTGTAT 368
Db 520 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
QY 369 AATGATGGTCTTCTGAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCTTTCGAG 488
Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
QY 489 CTTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAAATGAGAGTGAACAA 548
Db 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 619
QY 609 TCTTGGATCTCAGAGTCTCCTGAGACTGTTTTCACAGAAGCATGTGTGTACCCAG 668
Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
QY 669 GCTACACATCAAAAAGAAATGGAATAAATAAGTGAATAATAGAAATTCACATAGCCTA 728
Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
QY 729 TCAAAATCTTGGATACAGTTCATTCTTGTGAAGAGCAAGCAACTTCAAAAGATCAC 788
Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
QY 789 TGTGAACACGTACAGAAATAATGGAACAATGGAACAATGGAACAATGGAACAATGGAACA 848
Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
QY 849 AAATCTGTCAGAAGCAAAAGAAATAAATACAGTTAGAGAACCAAAAGTAAATGCGAA 908
Db 700 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 719
QY 909 CAAGAGCTCTCAGTGTGAGATTGACTTTAAACCAAGAACAGAGAGAGAGAGAGAGAGCC 968
Db 720 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 739
QY 969 GATATATTAAATGAAAAATTAGGGAAGAATTAGGAAGAATTCGAAGAGCAGCATAGAAA 1028
Db 740 AspileuAsnGluLysIleArgGluLeuGlyArgIleGluGluGlnHisArgLys 759
QY 1029 GAGTTAGAAGTGAACCAACACTTGAACAGGCTCTCAGATACAGATATAGATTGAAG 1088
Db 760 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 779
QY 1089 AGTGTAGAAATGAAATTTGAATCAGTTTCTCACACTCATCAAAATGAAATTTATCTCTTA 1148
Db 780 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 799
QY 1149 CATCAAAATTCATGTTTGAAGAAAGAAATTCGCCATGCTAAACTGGAATATAGCCACTG 1208
Db 800 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 819
QY 1209 AAACACCAATCCAGGAAAGAAATAAATACTTTGAGCAGCATTAGAGTTTAAAGAA 1268
Db 820 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 839
QY 1269 AAGATGCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAATCATTAACATAAAGGCA 1328

```
Db      840  LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 859
Qy      1329  TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACAAATGCTCACTTCTAAA 1388
Db      860  SerGlnTyrSerGlyGlnLeuLysValLeuAlaGluAsnThrMetLeuThrSerLys 879
Qy      1389  TTGAAGGAAAAACAAGACAAAGAAATACTAGAGCGAGAAATTCATCATCACCATCTCTAGA 1448
Db      880  LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 899
Qy      1449  CTGGCTTCTCTCTAGACAGACCATGATCAAAATGTGCATCAAGAAAAAGTCAAGAACCT 1508
Db      900  LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 919
Qy      1509  GCTTTCCACATTCGAGAGAGATGCTGTTGTCAAAGAAAAATGAATGTTGATGTAGTAGT 1568
Db      920  AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 939
Qy      1569  ACCGATATATAACAATGAGGTGCTCATCAACCACTTCTGAAGCTCAAGGAAATCCAN 1628
Db      940  Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 959
Qy      1629  AAGCTTAAATTAATCTCAATATATGACGAGATGCTCTAAGAGAAAAATACATTTGTTTC 1688
Db      959  sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuA:GluAsnThrLeuValSe 979
Qy      1689  AGAACATGCACAAAGAGACCAACGTGAACACAGTGTCAATGAGGAGCTGACACA 1748
Db      979  r-GluHisAlaGlnA:GaspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 999
Qy      1749  TGTNTCAANCGAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATC 1808
Db      999  etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuA:spG 1019
Qy      1809  AGAAATTAATTCACCTCAACAGCAAAATATGTGGCTTCAACAGCAATTAGTTCATGCA 1868
Db      1019  InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1039
Qy      1869  ATAANGAAGCTGACACAAACAAAGCAAGATAACAATTCATNTTCCTTGAGAGGAAA 1928
Db      1039  is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1058
Qy      1929  ATGCN-CATCATCTCTTAAAGAGAAAAATAGGAGATATTTNATACNATTAACATTATTA 1987
Db      1059  MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1078
Qy      1988  AAAAAACCGTATATTTCAATATGCAAAAAAANAANAANA 2027
Db      1079  LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1091

RESULT 13
ABJ37783
ID      ABJ37783 standard; protein; 1013 AA.
AC      ABJ37783;
XX      DT 15-MAY-2003 (first entry)
DE      Human tumour-related protein - SEQ ID No 553.
XX      DE
XX      Human; vaccine; gene therapy; T cell stimulation; T cell expansion;
XX      tumour; breast cancer; cancer; immune response stimulation.
XX      OS Homo sapiens.
XX      PN WO200283956-A1.
XX      PD 24-OCT-2002.
XX      PF 15-APR-2002; 2002WO-US012378.
XX      PR 13-APR-2001; 2001US-00834759.
XX      PR 07-DEC-2001; 2001US-00007805.
```

```
PR      13-FEB-2002; 2002US-00076622.
XX      (CORI-) CORIXA CORP.
XX      Houghton RL, Sleath PR, Persing DH, Jiang Y, Dillon DC;
XX      Mitcham JL, Xu J, Harlocker SL, Hepler WT, Henderson RA, Fanger GR;
XX      Vedvick TS, McNeill PD, Durham M;
XX      WPI; 2003-103376/09.
XX      New polypeptide and polynucleotide useful for stimulating and/or
XX      expanding T cells specific for a tumor protein and treating breast
XX      cancer.
XX      Example 8; Page 335-338; 375pp; English.
XX      The invention comprises a method of stimulating and/or expanding T cells
XX      specific for a tumour protein. The invention further comprises human
XX      nucleic acids and proteins that are associated with tumours (e.g. breast
XX      cancer). The method and sequences of the invention are useful for
XX      stimulating and/or expanding T cells specific for a tumour protein, a
XX      detecting the presence of cancer, stimulating an immune response in a
XX      patient and treating breast cancer. The present amino acid sequence
XX      represents a human tumour-related protein.
XX      SQ Sequence 1013 AA;

Alignment Scores:
Pred. No.:      5,51e-273      Length:      1013
Score:          3255.00      Matches:      652
Percent Similarity: 97.19%      Conservative: 4
Best Local Similarity: 96.59%      Mismatches: 15
Query Match:    95.65%      Indels:      5
DB:             6      Gaps:      0

US-09-602-362B-15 (1-2030) x ABJ37783 (1-1013)
```

```
Qy      12  AAAGATGCTTCTGAGGCTAACTGCGAATGAAAGTTTCTATTCCAACTAAAGCTTAA 71
Db      339  LysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAlaLeu 358
Qy      72  GAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAGAGGCAATCGCTTCGAG 131
Db      359  GluLeuMetAspMetGlnThrPheLysAlaGluProGlnLysProSerAlaPheGlu 378
Qy      132  CCTGCCATTCAAAATGCAAAAGTCTGTTCCAAAATAAAGCCTTGAATGAAGATGAACAA 191
Db      379  ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 398
Qy      192  ACATTGAGCAGATGAGATACCTCCCATCGAATCCAAACAAAGGACTATGAGAAAGT 251
Db      399  ThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGluSer 418
Qy      252  TCTTGGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTTGTGTTTACC 311
Db      419  SerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys 438
Qy      312  GCTACACATCAAAAAGAAATAGATAAAATAAATGGAATAATTAGAGAGTCTCTGATAAT 371
Db      439  AlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAspAsn 458
Qy      372  GATGGTTTCTGAGGCTCCCTCGACAATCAAGTTTCTATTCCCACTAAAGCCTTAGAA 431
Db      459  AspGlyPheLeuLysSerProCysArgMetLysValSerIleProThrLysAlaLeuGlu 478
Qy      432  TTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAGAGCCATCTCCCTCGAGCT 491
Db      479  LeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGluPro 498
Qy      492  GCCATTCAATGCAAAAGTCTGTTCCAAATAAGCCTTGAATGGAATGGAATGAACAAACA 551
Db      499  AlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGlnThr 518
```


QY 552 TTGAGACGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAATCT 611
 Db 519 LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysAsnValGluGluAsnSer 538
 QY 612 TGGGATCTCAGAGTCCGTCGAGACTGTTTCACAGAGGATGTTGTACCCAGGCT 671
 Db 539 TrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLysAla 558
 QY 672 ACACATCAAAAAGAAATGATAAATAAGTGGAAATAGAGATTCACATGACCTATCA 731
 Db 559 ThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeuSer 578
 QY 732 AAAATCTTCGATACAGTTCATCTCTGTAAGAGCAAGGAACTTCAAAAGATCACTGT 791
 Db 579 LysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHisCys 598
 QY 792 GAACAACTGACAGGAAAAATGAACAAATGAAAGAAAGTGTGTGCTACTGAAAAAGAA 851
 Db 599 GluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLysLys 618
 QY 852 CTGTCAGAGCAAGAAATAAATCACAGTTAGAGAACCAAAAGTTAAATGGGAACA 911
 Db 619 LeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGluGln 638
 QY 912 GAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAAATGCCGAT 971
 Db 639 GluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAlaAsp 658
 QY 972 ATATTAAATGAAATAATAGGAAAGAAATAGGAAAGATCGAAGACAGCATAGGAAAGAG 1031
 Db 659 IleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLysGlu 678
 QY 1032 TTAGAAGTGAACAACTTGAACAGGCTCTCAGATACAGATATAGAAATGAGAGT 1091
 Db 679 LeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLysSer 698
 QY 1092 GTAGAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATATCTCTTACAT 1151
 Db 699 ValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeuHis 718
 QY 1152 GAAATTTGATGTTGAAAGAAAGAAATGCGATGCTTAAACTGGAATAGCCACATGAAA 1211
 Db 719 GluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeuLys 738
 QY 1212 CACCAATACAGGAAAGAAATAAATACTTTGAGGACATTAAGATTTTAAAGAAAG 1271
 Db 739 HisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGluLys 758
 QY 1272 AATGCTGAATTCAGATGACCTTAAACTGAAAGAGGAATCAATTAATAAAGGCACTCT 1331
 Db 759 AsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAlaSer 778
 QY 1332 CAATATAGTGGCAGCTTAAAGTCTCTGATAGCTGAGAACACATGCTCACTTCTTAATG 1391
 Db 779 GlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLysLeu 798
 QY 1392 AAGGAAAAACAGACAAAGAAATCTAGAGGAGAAATGAATCAACACCATCTAGACTG 1451
 Db 799 LysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArgLeu 818
 QY 1452 GCTCTGCTGTACAGACCATGATCAATGTGACATCAAGAAAGATCAAGAACTGCT 1511
 Db 819 AlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluProAla 838
 QY 1512 TTCACATTCGAGGAGATGCTTTGTCAAAGAAATAATGAATGTTGATGTAGTAGTACC 1571
 Db 839 PheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerThr 858
 QY 1572 GATATATACATGAGGTGCTCATCAACACTTTCTGAAGCTCAAGGAAATCCANAAG 1631
 Db 859 -IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLysSe 878
 QY 1632 CCTTAAAAATTAATCTCAATTATGACGAGATGCTCTTAAGAGAAATACATTGGTTTCAGG 1691

Db 878 rLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer-G 898
 QY 1692 AACATGCACAAAGAGACCAACGTGAAACACAGTGTCAAATGAAGGAAGCTGAACACATGT 1751
 Db 898 luHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisMet 918
 QY 1752 NTCAANCGAACAGATNATGTGAACAAACACACTGANAGCAGGAGTCTCTAGATCAGA 1811
 Db 918 yrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspGlnL 938
 QY 1812 AATTATTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATAGTTTCATGCACATA 1871
 Db 938 ysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHis- 957
 QY 1872 ANGAAGCTGACCAACAAAGCAAAATCAATTAATTTTTCATTTCTTGAGAGGAAATG 1931
 Db 958 LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluA:GlysMet 977
 QY 1932 CN-CATCATCTCTTAAAGAGAAATGAGAGATTTTATTACNATACCACTTTTAAA 1990
 Db 978 GlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeuLys 997
 QY 1991 AACCCGTATATTTCAATATGAAAAAANAANAANAANA 2027
 Db 998 Asn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1009
 RESULT 14
 AAB84701
 ID AAB84701 standard; protein; 512 AA.
 XX
 AC AAB84701;
 DT 17-SEP-2001 (first entry)
 DE Amino acid sequence of a human cancer associated antigen.
 DE Cancer associated antigen; ING1; tumour suppressor; cancer; vaccine.
 KW Homo sapiens.
 OS
 XX
 FN WO2001.47959-A2.
 XX
 PD 05-JUL-2001.
 XX
 PF 29-NOV-2000; 2000WO-US042334.
 XX
 PR 30-NOV-1999; 39US-00451739.
 PR 24-OCT-2000; 2000US-00602362.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 PA (SLOK) SLOAN KETTERING INST CANCER RES.
 PA (CORR) CORNELL RES FOUND INC.
 XX
 PI Jager D, Stockert E, Scanlan M, Knuth A, Old L, Gure A, Chen Y;
 XX WPI; 2001-441706/47.
 DR N-PSDB; AAH28489.
 XX
 PT Isolated cancer associated nucleic acid molecule identified by SEREX
 PT (serological identification of antigens by recombinant expression
 PT cloning) technique, useful in nucleic acid based therapies to treat
 XX cancer.
 PS Claim 83; Page 50-51; 62pp; English.
 CC The present sequence represents a human cancer associated antigen. The
 CC gene is a tumour suppressor candidate gene. The cancer associated antigen
 CC polynucleotides and polypeptides are useful for screening for the
 CC possible presence of a pathological condition in a subject such as
 CC cancer. The cancer associated antigen polypeptides are useful for
 CC producing vaccines

XX	SQ	Sequence 512 AA;
Alignment Scores:		
Pred. No.:	3e-216	Length: 512
Score:	2599.00	Matches: 512
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	76.37%	Indels: 0
DB:	4	Gaps: 0
US-09-602-362E-15 (1-2030) x AAB84701 (1-512)		
Qy	42	ATGAAGATTCTATTCCCACTAAAGCCCTAGAAATGTGACATGCACAACTTTCAAGCA 101
Db	1	MetLysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAla 20
Qy	102	GAGCCTCCCGAGAAGCCATCTGCCTTCAGAGCTCCCATTTGAAATGCAAAAGTCTGTTC 161
Db	21	GiupProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerValPro 40
Qy	162	AATAAAGCCTTCGAATTGAAGAAATCAACAAACATTGAGACAGATGATACATCCCATCA 221
Db	41	AsnLysAlaLeuGluLeuLysAsnGlnThrLeuArgAlaAspGluIleLeuProSer 60
Qy	222	GAATCCAAACAAAGACATATGAAGAAAGTTCTTTGGGATCTTGAGAGTCTCTGTGAGCT 281
Db	61	GlusLysGlnLysAspTyrGluGluSerSerTrpAspSerGlusLysLeuLysGluThr 80
Qy	282	GTTTTCACAGAAGGATGTGTGTTTCCCAAGGCTACACATCAAAAAGAAATAGATAAAATA 341
Db	81	ValSerGlnLysAspValCysLeuProLysAlaThrHisGlnLysGluIleAspLysIle 100
Qy	342	AATGCAAAATTAGAAGAGTCTCCTGATTAATGATGGTTTTCTGAAGGCTCCCTCGAGAATG 401
Db	101	AsnGlyLysLeuGluGluSerProAspAsnAspGlyPheLeuLysAlaProCysArgMet 120
Qy	402	AAAGTTTCTATTCCAACTAAAGCCTTGAAATTTGATGGACATGCACAACTTTCAAAGCAGAG 461
Db	121	LysValSerIleProThrLysAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGlu 140
Qy	462	CTTCCGAGAAGCCATCTGCCTTCAGAGCTCGCATTTGAAATGCAAAAGTCTGTTCCAAT 521
Db	141	ProProGluLysProSerAlaPheGluProAlaIleGluMetGlnLysSerValProAsn 160
Qy	522	AAAGCCTTCGAATTGAAGAAATCAACAAACATTGAGACAGATCAGATGTTCCCTTCAGAA 581
Db	161	LysAlaLeuGluLeuLysAsnGluGlnThrLeuArgAlaAspGlnMetPheProSerGlu 180
Qy	582	TCAAAACAAAAGAGTTGAAGAAATTTCTTTGGGATCTTGAGAGTCTCCGTGAGACTGTT 641
Db	181	SerLysGlnLysLysValGluGluAsnSerTrpAspSerGlusLysLeuArgGluThrVal 200
Qy	642	TCACAGAAGGATGTGTGTGTCACCAAGGCTACACATCAAAAAGAAATGGATAAAATAAGT 701
Db	201	SerGlnLysAspValCysValProLysAlaThrHisGlnLysGluMetAspLysIleSer 220
Qy	702	GGAAATTAAGAAGATCAACTAGCCCTATCAAAATCTTTGGATACAGTTCACTTCTGTGAA 761
Db	221	GlyLysLeuGluAspSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGlu 240
Qy	762	AGAGCAAGGCACTTCAAAAAGATCACCTGTGCAACAAAGTACAGAAAAATGGACAACATG 821
Db	241	ArgAlaArgGluLeuGlnLysAspHisCysGluGluGlnArgThrGlyLysMetGluGlnMet 260
Qy	822	AAAAGAAGTTTGTGTACTGAAAAGAAACTGTCAGAACGCAAGAAATATAAATCACAG 881
Db	261	LysLysLysPheCysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln 280
Qy	882	TTTAGAGAAACAAAAAGTTTAAATGGGAAACAAAGCTCTGCAGTGTGAGATTGACTTTAAAC 941
Db	281	LeuGluAsnGlnLysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsn 300

```

PI Vedvick TS, Mcneill PD, Durham M;
XX WPI; 2003-103376/09.
XX New polypeptide and polynucleotide useful for stimulating and/or
XX expanding T cells specific for a tumor protein and treating breast
XX cancer.
XX Example 13; Page 359-363; 375pp; English.
XX The invention comprises a method of stimulating and/or expanding T cells
XX specific for a tumor protein. The invention further comprises human
XX nucleic acids and proteins that are associated with tumours (e.g. breast
XX cancer). The method and sequences of the invention are useful for
XX stimulating and/or expanding T cells specific for a tumour protein,
XX detecting the presence of cancer, stimulating an immune response in a
XX patient and treating breast cancer. The present amino acid sequence
XX represents a human tumour-related protein
XX Sequence 1239 AA;

Alignment Scores:
Pred. No.: 1,9e-186 Length: 1239
Score: 2256.50 Matches: 495
Percent Similarity: 62.54% Conservative: 61
Best Local Similarity: 55.68% Mismatches: 106
Query Match: 66.31% Indels: 227
DB: 6 Gaps: 6

US-09-602-362E-15 (1-2030) x ABU37789 (1-1239)
QY 21 CTTCTGAAGCTAACTGCGGAATGAAAGTTCTATTCCAACTAAAGCCTTAGAA---TTG 77
DB 343 LeupheGluSerSerAlaLysThrGlnValCysIleProGluSerMetTyrGlnLysVal 362
QY 78 ATGCACATGCAACTTTCAGACAGAGCCTCCGAGAGCCATCTGCTTCAGAGCTGCC 137
DB 363 MetGluIleAsnArgGluValGluGluLeuProGluLysProSerAlaPheLysProAla 382
QY 138 ATTGAATGCAAAAGTCTGTTCAMATAAAGCTTGAATTCGAAGTAATGAACAAACATG 197
DB 383 ValGluMetGlnLysThrValProAsnLysAlaPheGluLeuLysAsnGluGlnThrLeu 402
QY 198 AGAGCAGATGAGATCTCCATCAGAACTCAACAAAGAGACTATGAAGAAAGTCTCTGG 257
DB 403 ArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAspGluGluAsnSerTrp 422
QY 258 GATTCTGAGAGTCTCTGAGAGCTGTTTCACAGAGGATGTGTGTTTACCCAGGCTACA 317
DB 423 AspSerGluSerProCysGluThrValSerGlnLysAspValTyrLeuProLysAlaThr 442
QY 318 CATCAAAAAGAAATAGATAAAATAATGAAAATTAGAAAGTCTCTCTGATATGATGT 377
DB 443 HisGlnLysGluPheAspThrLeuSerGlyLysLeuGluGluSerProValLysAspGly 462
QY 378 TTCTCAGGCTCCCTGAGAAATGAAGTTCTATTCCAACTAAAGCCTTAGAATGATG 437
DB 463 LeuLeuLysProThrCysGlyArgLysValSerLeuProAsnLysAlaLeuGluLeuLys 482
QY 438 GACATCAAACTTTCAAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAGCCT----- 491
DB 483 AspArgGluThrPheLysAlaGluSerProAspLysAspGlyLeuLeuLysProThrCys 502
QY 492 -----GCCATTGAATGCCAAAG----- 509
DB 503 GlyArgLysValSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrLeuLys 522
QY 510 -----TCTGTT 515
DB 523 AlaGluSerProAspAsnAspGlyLeuLeuLysProThrCysGlyArgLysValSerLeu 542
QY 516 CCAAAATAAGCCTTGAAATGAAGATGAACAAACATTGAGAGCGAGATCAGATGTTCCCT 575
DB 598 GlnThrLeuArgIleGlnAspIleGluLeuLysSerValThrSerAsnLeuAsnGlnVal 917

```

```

543 ProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAlaAlaGlnMetPhePro 562
QY 576 TCAGAAATCAAAACAAAGAGTTGAAGAAATTTCTTGGGATTCAGAGTCTCCGTGAG 635
DB 582 SerGluSerLysGlnLysAspAspGluGluAsnSerTrpAspPheGluSerPheLeuGlu 582
QY 636 ACTGTTTCACAGAAGGATGTGTGTACCCCAAGGCTACACATCAAAAAGAAATCGATAAA 695
DB 593 ThrLeuLeuGlnAsnAspValCysLeuProLysAlaThrHisGlnLysGluPheAspThr 602
QY 696 ATAAGTGGAAATTTAGAA----- 713
DB 603 LeuSerGlyLysLeuGluGluSerProAspLysAspGlyLeuLeuLysProThrCysGly 622
QY 713 ----- 713
DB 623 MetLysIleSerLeuProAsnLysAlaLeuGluLeuLysAspArgGluThrPheLysAla 642
QY 713 ----- 713
DB 643 GluAspValSerSerValGluSerThrPheSerLeuPheGlyLysProThrThrGluAsn 662
QY 713 ----- 713
DB 663 SerGlnSerThrLysValGluGluAspPheAsnLeuThrThrLysGluGlyAlaThrLys 682
QY 713 ----- 713
DB 683 ThrValThrGlyGlnGlnGluArgAspIleGlyIleIleGluArgAlaProGlnAspGln 702
QY 713 ----- 713
DB 703 ThrAsnLysMetProThrSerGluLeuGlyArgLysGluAspThrLysSerThrSerAsp 722
QY 713 ----- 713
DB 723 SerGluIleIleSerValSerAspThrGlnAsnTyrGluCysLeuProGluAlaThrTyr 742
QY 713 ----- 713
DB 743 GlnLysGluIleLysThrThrAsnGlyLysIleGluGluSerProGluLysProSerHis 762
QY 713 ----- 713
DB 763 PheGluProAlaThrGluMetGlnAsnSerValProAsnLysGlyLeuGluTrpLysAsn 782
QY 714 -----GATTCAACTAGCCTATCAAAAATCTTTGGATACAGTTTCATTCT 755
DB 783 LysGlnThrLeuArgAlaAspSerThrThrLeuSerLysIleLeuAspAlaLeuProSer 802
QY 756 TGTGAAAGAGCAAGGAACTTCAAAAAGATCACTGTGAACACACGTCAGGAAAAATCGAA 815
DB 803 CysGluArgGlyArgGluLeuLysLysAspAsnCysGluGlnIleThrAlaLysMetGlu 822
QY 816 CAAATGAAAAGAGATTTGTGTACTGAAAAGAACTGTCAAGACAAAAGAAATAAAA 875
DB 823 GlnMetLysAsnLysPheCysValLeuGlnLysGluLeuSerGluAlaLysGluLys 842
QY 876 TCAAGTTAGAACACAAAAGTTAAATGGAAACAGAGCTCTGCACTGTGAGATTGACT 935
DB 843 SerGlnLeuGluAsnGlnLysAlaLysTrpGluGlnGluLeuCysSerValArgLeuPro 862
QY 936 TTAACCAAGAGAGAGAGAGAGAAATGCCGATATATTAAATGAAAATATAGGAA 995
DB 863 LeuAsnGlnGluGluGluLysArgAsnValAspIleLeuLysGluLysIleArgPro 882
QY 996 GAATTAGGAAGATTCGAGAGCAGCATAGGAAGAGTTAGAAAGTAGAAACAACTTGAA 1055
DB 883 -----GluGluGlnLeuArgLysLysLeuGluValLysHisGlnLeuGlu 897
QY 1056 CAGCTCTCAGATACAGATATAGAAATTCAGAGTCTAGAAAGTAAATTTGAATCAGTT 1115
DB 898 GlnThrLeuArgIleGlnAspIleGluLeuLysSerValThrSerAsnLeuAsnGlnVal 917

```

```
Qy 1116 TCTCACACTCATGAAATGAAATTTATCTCTTACATGAAATTTGCATGTTGAAAGGAA 1175
Db 918 SerHisThrHisGluSerGluAsnAspLeuPheHisGluAsnCysMetLeuLysLysGlu 937
Qy 1176 ATTGCCATGTAAATGGAATAGGCACACTGAAACACCAATACCAGGAAAGGAAAT 1235
Db 938 IleAlaMetLeuLysLeuGluValAlaThrLeuLysHisGlnHisGlnValLysGluAsn 957
Qy 1236 AATACATTTGAGGACATTAAGATTTTAAAGAAAGAAATGCTGAACTTCAGATGACCTA 1295
Db 958 LysTyrPheGluAspIleLysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeu 977
Qy 1296 AATGTGAAGAGGAATCATTAACATAAAGGGCATCTCAATATATAGTGGCAGCTTAAAGTT 1355
Db 978 LysLeuLysGlnLysThrValThrLysArgAlaSerGlnTyrArgGlnGlnLeuLysVal 997
Qy 1356 CTGATAGCTGAGAACACATGCTCCTTCTTAAATTTGAAGGAAACCAAGACAAAGAAATA 1415
Db 998 LeuThrAlaGluAsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIle 1017
Qy 1416 CTAGAGCCAGAAATGAAATGACACCATCTCTGACTGCTGTCTGTACAGACCATGAT 1475
Db 1018 LeuGlnThrGluIleGluSerHisHisProArgLeuAlaSerAlaLeuGlnAspHisAsp 1037
Qy 1476 CAAATTTGTGACATCAAGAAAGTCAAGAACCTGCTTCCACATTTGCAGGAGATGCTTGT 1535
Db 1038 GlnSerValThrSerArgLysAsnGlnGluLeuAlaPheHisSerAlaGlyAspAlaPro 1057
Qy 1536 TTGCAAGAGAAATGAAATGTTGTGAGTAGTACCGATATATAACAATGAGGTGCTCCA 1595
Db 1058 LeuGlnGlyIleMetAsnValAspValSerAsnThr-IleTyrAsnAsnGluValLeuHi 1077
Qy 1596 TCACACACTTCTGAACTCAAGGAAATCCANAAGCTTAAATAATTAATCTCAATATTATGC 1655
Db 1077 SGlnProLeuTyrGluAlaGlnArgLysSerLysSerProLysIleAsnLeuAsnTyrAl 1097
Qy 1656 AGGAGATGCTCTAAGAGAAATACATGTTTTCAGGAACATGACACAAAGAGACCAACGTG 1715
Db 1097 aglyAspAspLeuArgGluAsnAlaLeuValSer-GluHisAlaGlnArgAspArgCysG 1117
Qy 1716 AATCACAGTGTCAATGAAGAGAGCTGAACACATGTTCAAACGAAACAAAGATNATGTGA 1775
Db 1117 IuThrGlnCysGlnMetLysLysAlaGluHisMetTyrGlnAsnGluGlnAspAsnVala 1137
Qy 1776 ACAACACACTGANCACGAGTCTCTAGATCAGAAATTTATTCAACTACAAAGCAAAA 1835
Db 1137 spLysHisThrGluGlnGlnGluSerLeuGlnLysLeuPheGlnLeuGluSerLysA 1157
Qy 1836 ATATGCTGCTTCAACAGCAATTAAGTTCATGCAATAANGAAGCTGACACACAAAGCAAG 1895
Db 1157 snArgTrpLeuArgGlnGlnLeuValTyrAlaHis-LysLysVal--AsnLysSerLys 1175
Qy 1896 ATAACAATTTGATTTTCTTTGAGAGGAAATGCNATCATCTTCTTAAAGAGAAAA 1955
Db 1176 ValThrIleAsnIleGlnPheProGluMetLysMet--GlnArgHisLeuLysGluLysA 1195
Qy 1956 ATGAGGAGATATTNATACNATACCATTTAAACCCGATATTTCAATATGGAATA 2015
Db 1195 snGluGluValPheAsnTyrGlyAsnHisLeuLysGlu-ArgIleAspGlnTyrGluLys 1214
Qy 2016 AAAAAAANAANAANA 2030
Db 1215 GluLysAlaGluArg 1219
```

Search completed: July 15, 2004, 09:05:46
Job time : 149.357 secs

GenCore version 5.1.6
 Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:46:47 ; Search time 95.3178 Seconds
 (without alignments)
 13439.296 Million cell updates/sec

Title: US-09-602-362E-15
 Perfect score: 3403
 Sequence: 1 ctgcggcgttaaatggt.....gaaaaaaaaaaaaaaaaa 2030

Scoring table: BLOSUMP62
 Xgapop 10.0 , Xgapext 0.5
 Ygapop 10.0 , Ygapext 0.5
 Fgapop 6.0 , Fgapext 7.0
 Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+ n2p model -DEV=xlp
 -Q/cgn2_1/USPTO_spool_p/US09602362/runat_15072004_093626_22001/app_query.fasta_1.10325
 -DB=SPREMBL_25 -QFMT=Fascan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
 -UNITS=bits -START=1 -END=1 -MATRIX=blosump62 -TRANS=human40.cdi -LIST=45
 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
 -OUTFMT=ptc -NORM=ext -HEAPSZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09602362 @CGN 1 1 814 @runat 15072004_093626_22001 -NCPU=6 -ICPU=3
 -NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
 -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	3272	96.2	1341 4	Q9bxx3 homo sapien

2	1559.5	45.8	424	4	Q9NSI9	Q9nsi9 homo sapien
3	1546	45.4	1011	4	Q9BXX2	Q9bxx2 homo sapien
4	1190.5	35.0	1709	4	Q9UPF8	Q9upf8 homo sapien
5	1188	34.9	1710	4	Q9HIQ1	Q9hiq1 homo sapien
6	790.5	23.2	718	4	Q9HCD1	Q9hcd1 homo sapien
7	671.5	19.7	823	4	Q9H0H6	Q9h0h6 homo sapien
8	665.5	19.5	641	4	O60311	O60311 homo sapien
9	621	18.2	453	4	Q8IZM7	Q8izm7 homo sapien
10	443.5	13.0	996	4	Q8IVF6	Q8ivf6 homo sapien
11	330	9.7	2055	5	Q8TSC7	Q8tsc7 plasmodium
12	330	9.7	2055	5	Q8IHP3	Q8ihp3 plasmodium
13	318.5	9.4	1738	5	O76329	O76329 dictyosteli
14	314	9.2	520	11	Q8OW27	Q8ow27 mus musculu
15	302.5	8.9	381	4	Q15694	Q15694 homo sapien
16	285	8.4	2011	5	Q8MMC2	Q8mmc2 drosophila
17	285	8.4	2056	5	Q9W0W8	Q9w0w8 drosophila
18	280	8.2	2017	5	Q94992	Q94992 drosophila
19	280	8.2	2057	5	Q94987	Q94987 drosophila
20	272	8.0	1305	10	Q9FJ35	Q9fj35 arabidopsis
21	269	7.9	2033	10	Q7XEH4	Q7xeh4 oryza sativ
22	267	7.8	1979	5	O96133	O96133 plasmodium
23	267	7.8	2954	13	O42263	O42263 xenopus lae
24	284.5	7.8	1313	4	Q9HCK7	Q9hck7 homo sapien
25	264.5	7.8	1831	10	Q7XN11	Q7xnl1 oryza sativ
26	262	7.7	1388	6	Q28021	Q28021 bos taurus
27	261	7.7	2007	13	Q02015	Q02015 gallus gall
28	257.5	7.6	1320	11	Q9JK25	Q9jk25 ratus norv
29	257	7.6	1208	5	Q9VXU1	Q9vxu1 drosophila
30	257	7.6	1398	5	Q9VXU2	Q9vxu2 drosophila
31	256.5	7.5	895	11	O81112	O81112 mus musculu
32	256.5	7.5	1790	3	Q07380	Q07380 saccharomyc
33	256.5	7.5	1960	11	Q8VDD5	Q8vdd5 mus musculu
34	255.5	7.5	1762	10	Q94DC2	Q94dc2 oryza sativ
35	255	7.5	1938	6	Q9GJP9	Q9gjp9 cryctolagus
36	255	7.5	2101	4	Q14981	Q14981 homo sapien
37	255	7.5	2115	4	Q14980	Q14980 homo sapien
38	254	7.5	2245	5	Q86A36	Q86a36 dictyostell
39	253	7.4	1416	4	Q9BZF9	Q9bzf9 homo sapien
40	253	7.4	1416	4	Q9HCL1	Q9hcl1 homo sapien
41	253	7.4	1449	6	Q9BG87	Q9bg87 bos taurus
42	253	7.4	2253	13	P70012	P70012 xenopus lae
43	252	7.4	947	5	Q8ICA7	Q8ica7 plasmodium
44	252	7.4	1388	6	Q9GL21	Q9gl21 canis famil
45	251	7.4	2029	4	Q9C014	Q9c014 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9BXX3	PRELIMINARY;	PRT;	1341 AA.
AC	Q9BXX3;			
DT	01-JUN-2001 (TRENBLrel. 17, Created)			
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)			
DT	01-OCT-2003 (TRENBLrel. 25, Last annotation update)			
DE	Breast cancer antigen NY-BR-1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21174979; PubMed=11280766;			
RA	Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karch J.,			
RA	Jaeger E., Knuth A., Old L.J., Chen Y.T.;			
RT	"Identification of a tissue-specific Putative Transcription Factor in			
RT	Breast Tissue by Serological Screening of a Breast Cancer Library."			
RL	Cancer Res. 61:2055-2061(2001).			
DR	EMBL; AF269087; AAK27325.1; -			
DR	GO; GO:0005634; C:nucleus; NAS.			
DR	GO; GO:0005515; F:protein binding; NAS.			
DR	GO; GO:0003700; F:transcription factor activity; NAS.			
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.			

DR InterPro; IPR002110; ANK.
 DR InterPro; IPR001969; Aspprotease_AS.
 DR Pfam; PF00023; ank; 6
 DR SMART; SM00248; ANK; 6.
 DR PROSITE; PS50088; ANK REPEAT; 4.
 DR PROSITE; PS50297; ANK REP REGION; 1.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 KW ANK repeat; Repeat.
 SQ SEQUENCE 1341 AA; 152777 MW; 33E53DDBE6FD3A58B CRC64;

Alignment Scores:
 Pred. No.: 9,34e-192 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: Gaps: 0

US-09-602-362E-15 (1-2030) x Q9BXK3 (1-1341)

QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
 DB 666 VALLyAspGLyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
 QY 69 TTGAATTGATGACATGCAAACTTTCAAGACGAGCCTCCGAGAGCCATCTCGCTTC 128
 DB 686 LeuGLuLeuMetAspMetGlnThrPheLysAlaGluProProGLuLysProSerAlaPhe 705
 QY 129 GAGCTGCCATTCAATGCAAAAGTCTGTTCAAAATAAAGCCCTTGGAAATTGAAGAAATGAA 188
 DB 706 GLUProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
 QY 189 CAAACATTGACGACATGAGATCTCCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
 DB 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTTCTTGGATTCAGAGTCTCTGTGAGACTGTTTCACAGAAGGATGTGTTTACCC 308
 DB 746 SerSerTrpAspSerGluSerIeuCysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 AAGCTACACATCAAAAGAAATPAGATAAAATAAATGGAATAATTAGAGAGTCTCTGAT 368
 DB 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLysSerProAsp 785
 QY 369 AATGATGTTTCTGAAGCTCCCTCGAGAATCAAGATTTCTATCCCACTAAAGCCTTA 428
 DB 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCCGAGAAGCCATCTCGCTTCGAG 488
 DB 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
 QY 489 CTGCGCATGGAATGCAAAAGCTGTTCMAATAAAGCCTTGGAAATGGAATGGAACAA 548
 DB 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACGACATCAGATGTTCCTCTCAGAAATCAAAACAAAGAGTTTCAAGAAAT 608
 DB 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGGATTCGAGAGTCTCGTGAGACTGTTTCAGAGAAGGATGTGTGTACCCAG 668
 DB 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGGAATAAAATGGAATAAATGGAATCAACTAGCCCTTA 728
 DB 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
 QY 729 TCAAAATCTTGATACAGTTCTTCTGTGAAGACGAGGAACTTCAAAAGATCAC 788
 DB 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
 QY 789 TGTGAACAACGTACAGGAAATGGAACAAATGAAAGAAAGAGTCTTGTGTACTGAAAGAA 848

DB 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
 QY 849 AAAGTGTGAGAGCAGAAAGAAATAAATACAGTTAGAGAACCAAAAGTTAAATGGAA 908
 DB 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGlnLysValLysTyrGlu 965
 QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTTAAACCAAGAGAGAGAGAGAAATGCC 968
 DB 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluGluLysArgAlaGlnAla 985
 QY 969 GATATATTAATGAATAATTAGGAGAAATTAGGAGAAATAGGAGAAATAGGAGAGAGAG 1028
 DB 986 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGlnGlnHisArgLys 1005
 QY 1029 GAGTTAGAGTGAACCAACCAACTTGAACAGCTCTCAGATACAGATATAGAAATGAAG 1088
 DB 1006 GluLeuGluValLysGlnGlnLeuGluAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTGAAGATTAATTGAATCAGTTTCTCACATCATGAAATGAAATATATCTCTTA 1148
 DB 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
 QY 1149 CATGAAATATGCAATGTTGAAAAAGAAATTCCTCATGCTAAACTGGAATAGCACACTG 1208
 DB 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AAACACCAATACCAAGGAAAGAAATAATACTTTGAGACATTAAGATTTTAAAGAA 1268
 DB 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAACTCTCAACTTCAGATGACCTAAACTCAAAGAGGAATCATTAACATAAAGGCA 1328
 DB 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLeuSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACCAATGTCTACTTAA 1388
 DB 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAGGAAAAACAGCAAGAAATACTAGAGGCGAATTCGAATCACACCATCCTAGA 1448
 DB 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 1145
 QY 1449 CTGCTTCTCTGTGTACAGACCATGATCAAAATGTGACATCAAGAAAAAGTCAAGAACCT 1508
 DB 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCACATTCGAGAGATGCTTGTTCAGAGAAAAATGAATGTTGATGTGAGTAGT 1568
 DB 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATAACAAATGAGGTGCTCCATCAACACTTCTGAAAGCTCAAGAGAAATCCAN 1628
 DB 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
 QY 1629 AAGCTTAAATTAATCTCAATTATGAGAGATGCTCTTAAGAGAAAAATACATGTGTTTC 1688
 DB 1205 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
 QY 1689 AGGAACATGCACAAAAGAGACCAACGTCGAAACACAGTGTCAAAATGAAGAGAGCTGAAACACA 1748
 DB 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
 QY 1749 TGTNTCAAAACGACAGATNATGTGAACAAAACACTGANCAGCAGGAGTCTCTAGATC 1808
 DB 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAATTTATTCATCTACAAAGCAAAATATGCGTTCACACAGCAATAGTTTCATGCAC 1868
 DB 1265 LnyLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaH 1285
 QY 1869 ATAANGAAAGCTGACAAACAAAGCAAGATAACTCATNTTCATNTCTTGAGAGGAAA 1928

```
Db 1285 is-LysLysAlaAspAsnLysSerLysThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ARGON-CATCATCTCTAAAGAGAAAATAGAGAGATATTNTATCAATCAATCAATTA 1987
Db 1305 MetGlnHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAACCCGTATATTCAATATGGAAGAAAAGAAAAGAAA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyr-GluLysGluLysAlaGlu 1337

RESULT 2
QNSI9
ID QNSI9 PRELIMINARY; PRT; 424 AA.
AC QNSI9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PRE4 protein (Fragment).
GN PRE4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald A., Kump A., Schillabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mituyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Rieseemann L., Dagand E.,
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lehrach H., Reinhardt R., Vaspo M.L.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL163203; CAB90394.1; -.
FT NON_TER 1
FT NON_TER 424
SQ SEQUENCE 424 AA; 50038 MW; C6E704795C534625 CRC64;

Alignment Scores:
Pred. No.: Length: 424
Score: 1559.50 Matches: 334
Percent Similarity: 82.84% Conservative: 33
Best Local Similarity: 75.40% Mismatches: 55
Query Match: 45.83% Indels: 21
DB: Gaps: 3

US-09-602-362E-15 (1-2030) x QNSI9 (1-424)
QY 714 GATTCACTAGCTATCAAAAATCTGGATACAGTTCATTCTTGAAAGAGCAAGGAA 773
Db 2 AspSerThrThrLeuSerLysLeuLeuAlaValProSerCysGluArgGlyLysGlu 21
QY 774 CTTCAAAAAGATCATCTGCAACACCTACAGGAAAATGGAACAAATGAAAAGAGTTT 833
Db 22 LeuLysLysAspHisCysGluGlnIleThrAlaLysMetGluGlnThrLysAsnLysPhe 41
QY 834 TGTGTACTGAAAAGAACTGTGAGAGCAAGCAAGAAAATAAATACAGTTAGAGAACCA 893
Db 42 CysValLeuGlnLysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGln 61
QY 894 AAAGTTAAATGGAAACAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAG 953
Db 62 LysAlaLysTrpGluGlnLeuLysCysSer-----Lys 72
QY 954 AAGAGAGAAGATGCCATATATTAATGAAAATTAGGAAGAAATTAGGAAGATCGAA 1013
Db 73 LysArgArgAspValAspIleLeuLysGluLysIleArgPro-----Glu 87
```

```
QY 1014 GAGCGCATAGGAAGAGTTAGAGTGAAACAACTTGAACAGGCTCTCAGAATACAA 1073
Db 88 GluGlnLeuArgLysLysLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgLysGln 107
QY 1074 GATATAGAATTGAAGAGTGTAGAAAGTAATTTCAATCAGGTTTCTCACACTCATGAAT 1133
Db 108 AspIleGluLeuLysSerValThrSerAsnLeuAsnGlnValSerHisThrHisGluSer 127
QY 1134 GAAAATTATCTCTACATGAAATTCATGTTCAAAAGGAAATTCGCATGCTAAAGCTG 1193
Db 128 GluAsnAspLeuPheHisGluLysCysMetLeuLysGluIleAlaMetLeuLysLeu 147
QY 1194 GAAATAGCCACACTGAAACACCAATACAGGAAAAGGAAAATAAATACATTTTGAGGCA 1253
Db 148 GluValAlaThrLeuLysArgGlnHisGlnValLysGluAsnLysTyrPheGluAspIle 167
QY 1254 AAGATTTTAAAGAAAAGATGCTGAACITTCAGATGACCCCTAAACTGAAGAGGAATCA 1313
Db 168 LysIleLeuGlnGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGlnLysThr 187
QY 1314 TTAACATAAAGGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACA 1373
Db 188 LeuThrLysArgAlaSerGlnTyrArgGluGlnLeuLysValLeuThrAlaGluAsnThr 207
QY 1374 ATGCTCACTTCTAAATTGAAGGAAAACAAAGACAAAGAAATACCTAGCGGAGAAATGAA 1433
Db 208 MetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGluThrGluIleGlu 227
QY 1434 TCACACCATCTAGACTGGCTTCTGCTGTACAAGACCATCATCAAAATTTGTGACATCAAGA 1493
Db 228 SerHisHisProArgLeuAlaSerAlaLeuGlnAspHisAspGlnSerValThrSerArg 247
QY 1494 AAAAGTCAAGAACCTGTCTTCCACATTCGAGGAGATCTTGTGTTGCAAGAAAATAATGA 1553
Db 248 LysAsnGlnGluLeuAlaPheHisSerAlaGlyAspAlaHisLeuGlnGlyIleMetAsp 267
QY 1554 GTTCATGTGAGTAGTACCGATATATAACATGAGCTCTCCATCAACCATCTTCTGAGC 1613
Db 268 ValAspValSerAsnThr-IleTyrAsnAsnGluValLeuHisGlnProLeuTyrGluAl 287
QY 1614 TCAAGGAAAATCCANAAGCCCTAAAAATTAATCTCAATTATGCGAGGAGATCTCTTAAGAG 1673
Db 287 aGlnArgLysSerLysSerProLysIleAsnLeuAsnTyrAlaGlyAspAspLeuArgG 307
QY 1674 AATACATCTGTTTTCAGGACATGCAACAAAGAGACCAACCTGAAACACACTGTCAATGA 1733
Db 307 uAsnAlaLeuValSer-GluHisAlaGlnArgAspArgCysGluThrGlnCysGlnMetL 327
QY 1734 AGGAAGTGAACACATCTNTCAAACGAAACAAAGATNATGTGAACAAACACACTGANCAGC 1793
Db 327 yLysAlaGluHisMetTyr-GlnAsnGluGlnAspAsnValAspLysHisThrGluGlnG 347
QY 1794 AGGAGTCTTAGATCAGAAATATTTCACACTACAAGCAAAAATATGTGGCTTCAACAGC 1853
Db 347 InGluSerLeuGluGlnLysLeuPheLysLeuGluSerLysAsnArgTrpLeuArgGlnG 367
QY 1854 AATTAGTTCATGCAATAANGAAGCTGCAACAAGAAAGCAAGATAACAATTTGATTTTCAT 1913
Db 367 InLeuValTyrAlaHis-LysLysVal---AsnLysSerLysValThrIleAsnIleGln 385
QY 1914 TMTCTTGAGAGGAAAATGCNCATCATCTTCTAAAGAGAAAATGAGGAGATATTNATT 1973
Db 386 PheProGluThrLysMet--GlnArgHisLeuLysGluLysAsnGluValPheAsnT 405
QY 1974 ACNATAACCATTTTAAAAACCCGTTATTTCAATATGGAAGAAAAGAAAAGAAAAGAAA 2030
Db 405 yrGlyAsnHisLeuLysGlu-CysIleAspGlnTyrGluLysGluLysAlaGluArg 423

RESULT 3
QNSI9
ID QNSI9 PRELIMINARY; PRT; 1011 AA.
AC QNSI9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
```



```

QY 1115 ----- 1115
Db 859 AspAlaGlnArgInLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyIleLeu 878
QY 1115 ----- 1115
Db 879 ThrAsnHisLeuSerLysGlnLysGluIleGluMetAlaGlnLysLysMetAsnSerGlu 898
QY 1116 ---TCTCACACTCATGAATGAAATATATCTTACATGAAATGCGATGTTGAAAG 1172
Db 899 AsnSerHisSerHisGluGluGluLysAspLeuSerHisLysAsnSerMetLeuGlnGlu 918
QY 1173 GAAATGCGCATGCTAAACTGGAATAGCCACACTGAAACCAATACCCAGGAAAGGAA 1232
Db 919 GluIleAlaMetLeuArgLeuGluIleAspThrIleLysAsnGlnAsnGlnGluLysGlu 938
QY 1233 AATAATATCTTGGAGACATTAAGATTTAAAGAAAGAAAGATGCTCAACTTCAGATGACC 1292
Db 939 LysLysCysPheGluAspLeuLysIleValLysGluLysAsnGluAspLeuGlnLysThr 958
QY 1293 CTAAACTGAAAGAGGAATCATTAATAAAGGCGCATCTCAATATAGTGGCAGCTTAA 1352
Db 959 IleLysGlnAsnGlnGluThrLeuThrGlnThrIleSerGlnTyraSngLysLeuSer 978
QY 1353 GTTCTGATAGTACAGACATGCTCACTTCTTAATTTGAAG---GAAAAACAAGACAA 1409
Db 979 ValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGlnLysSerLys 998
QY 1410 GAAATATCTAGAGGACAGAAATTTGAATCACACCATCTCTGCTCTCTCTCAAGAC 1469
Db 999 GluArgLeuGluAlaGluValGluSerTyHisSerArgLeuAlaAlaIleHisasp 1018
QY 1470 CATGATCAAAATTTGATCATCAAGAAAGAACTCAAGAACCTGCTTCCATTCAGAGAGAT 1529
Db 1019 ArgAspGlnSerGluThrSerLysArgGluLeuLeuAlaPheGlnArgAlaArgasp 1038
QY 1530 GCTTGT-----TTGCAAGAAAGAAATGAATGTGATGTGAGTAGTACCGATATACAA 1583
Db 1039 GluCysSerArgLeuGlnAspLysMetAsnPheAspValSerAsnLeuLys-AspAsn 1058
QY 1584 TGAGGTGCTCCATCAACCACTTCTGAAGCTCAAGGAAATCCANAAGCTTAAATTTAA 1643
Db 1058 nGluIleLeuSerGlnGlnPheLysThrGluSerLysLeuAsnSerLeuGluIleG 1078
QY 1644 TCTCAATATGAGGAGAGCTTAAGAGAAATACATGTTGTTTCCAGGACATGCACAA 1703
Db 1078 uPheHisHisThrArgAspAlaLeuArgGluLysThrLeu-GlyLeuGluArgValGlnL 1098
QY 1704 GAGACAACTGAAACACACAGTGTCAATGAAGGAACTGAAACATGTTCAAAACGAAAC 1763
Db 1098 ysAsp-leuSerGlnThrGlnCysGlnMetLysGluMetGluGlnLysTyrglnAsnGluG 1118
QY 1764 AGATNATGTGAACAAACACATGANCACAGAGTCTTAGATCAGAAATATTATTTCAAC 1823
Db 1118 InValLysValAsnLysTyriIleGlyLysGlnGluSerValGluGluArgLeuSerGlnL 1138
QY 1824 TACAAAGCAAAATATGTGCTTCAACAGCAATATTAGTTTCATGCACATAANGAAAGCTCAC 1883
Db 1138 euGlnSerGluAsnMetLeuLeuArgGlnGlnLeuAspAlaHis-AsnLysAlaasp 1157
QY 1884 ACAAAAGCAAGATACAAATGAT-----NTTCATNTTCTTGAGAGGAAATG 1931
Db 1158 AsnLysGluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLysLeu 1177
QY 1932 CNCATCAT-----CTTCAAGAGAGAAATGAGGAGATATTTNAT 1972
Db 1178 GlnAlaGluSerGluLysGlnSerLeuLeuLeuGluGluArgAsnLysGluLeuIleSer 1197
QY 1973 TACNATAACATTTAAAAACCCGTATATTTCAATATGGAATAAAAAAANAATAAAAA 2030
Db 1198 GluCysAsnHisLeuLysGlu-ArgGlnTyrglnTyrglnGluAsnGluLysAlaGluArg 1216

```

```

RESULT 5
Q9H1Q1
AC Q9H1Q1 PRELIMINARY; PRT; 1710 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE BAI45E8.1 (Hypothetical protein KIAA1074).
GN BAI45E8.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162272; CAC19649.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 4.
DR SMART; SM00248; ANK; 4.
DR PROSITE; PS00086; ANK REPEAT; 4.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein; ANK repeat; Repeat.
SQ SEQUENCE 1710 AA, 196410 MW, 01CBF9BADB894872 CRC64;

```

Alignment Scores:

```

Pred. No.: 1e-64 Length: 1710
Score: 1188.00 Matches: 309
Percent Similarity: 56.62% Conservative: 123
Best Local Similarity: 40.50% Mismatches: 213
Query Match: 34.91% Indels: 119
DB: 4 Gaps: 16

```

US-09-602-362E-15 (1-2030) x Q9H1Q1 (1-1710)

```

QY 33 AACTGCGGAATGAAAGTTTCTTCTTCCAACTAAAGCCTTAGAATTGTGACATGCAAACT 92
Db 479 AsnValGlyMetProValAla----- 485
QY 93 TTCAAACGACAGCGCTCCGAGAAAGCCATTCGCTTCGAGCTGCCATTGAAATGCAAAAG 152
Db 486 ---HisMetGluSerProGluArgTyLeuHisLeuLysProThrIleGluMetLysAsp 504
QY 153 TCTGTTCCAAATAAGCTTGGAAATTCAGAAATGAACAACATTCGAGACAGATGAG--- 209
Db 505 SerValProAsnLysAlaGlyGlyMetLysAspValGlnThrSerLysAlaAlaGluHis 524
QY 210 ATACTCCCATCAGAATCCAAACAAAGGACTATGAAGAAAGTTCTTGGGATTCGAGAGT 269
Db 525 AspLeuGluValAlaSerGluGluGluGlnGluArgGluGlySerGluAsnAsnGlnPro 544
QY 270 CTCTGTGAGACTTTTCACAGAGAGATGTGTGTTTACCACAGGCTACACATCAAAAGAA 329
Db 545 GlnValGluGluArgLysLys-----HisArgAsnAsnGluMetGlu 559
QY 330 ATA-----GATAAATAAATGGAATAATTAGAAAGAGTCTCTCTGATAATGAT 374
Db 560 ValSerAlaAsnIleHisaspGlyAlaThrAspAspAlaGluAspAspAspAsp 579
QY 375 GGTGTTTCTG-----AAGGCTCCCTCGAGNATGAAAGTTTCTTATTCCACTAAA--- 422
Db 580 GlyLeuIleGlnLysArgLysSerGlyThrAspHisGlnGlnPheProArgLysGlu 599
QY 423 -----GCTTAGAATTGATGCATGCAAA---ACTTTC 452
Db 600 AsnLysGluTyraLaserSerGlyProAlaLeuGlnMetLysGluValLysSerThrGlu 619
QY 453 AAGCAGAGCTCCCGAGAGCCATTCGCTTCGAGCCTGCCATTCGAAATGCAAAAGTCT 512
Db 620 LysGluLysArgThrSerLysGluSerValAsnSerProValPheGlyLysAlaSerLeu 639
QY 513 GTTCCAAATAAAGCCTTGGAAATTGAAGAAATGAACAACATG----- 554

```

```

Db 640 LeuThrGlyGlyLeuLeuGlnValAspAspSerSerLeuSerGluLeuAspGluAsp 659
QY 555 -----AGAGCAGATCAGATGCTCCCTTCAGAAACAAACAAAGAGAGGTTGAAGAAAAT 608
Db 660 GluGlyArgProThrLysLysThrSerAsnGluLysAsnValLysAsnGlnLeuGln 679
QY 609 TCTTGGGATTCT-----GAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTG 656
Db 680 SerMetAspAspValAspAspLeuThrGlnSerSerGluThrAlaSerGlu---AspCys 698
QY 657 TGTGTACCAACAGGCTACACATCAAAAGAAATG-----GATAAAATAAGTCGAAAA 707
Db 699 GluLeuProHisSerSerTyLysAsnPheMetLeuLeuLeuGluGlnLeuGlyMetGlu 718
QY 708 TTAGAAGATTCAACTAGGCTATCAAAATCTTGATGATCAGGTTCAATCTTGTTGAAGAGCA 767
Db 719 CysLysAspSerValSerLeuLysIleGlnAspAlaLeuSerCysGluGluLeu 738
QY 768 AGGGAACCTTCAAAAGATCACTGTGAACACAGTACAGGAAAAATGGACAAATCAAAAG 827
Db 739 LeuGluLeuLysLysAsnHisCysGluLeuLeuThrValLysIleLysLysMetGluAsp 758
QY 828 AAGTTTGTGTACTGAAAAAGAACTGTGAGAAGCAAAAGAAATAAATCAATCAGTTAGAG 887
Db 759 LysValAsnValLeuGlnArgGluLeuSerGluThrLysGluLeuLysSerGluLeuGlu 778
QY 888 AACCAAAAGATTAAATGGGAACAGAGCTCTGCAGTGTGAGATTGACATTTAAACCAAGAA 947
Db 779 HisGlnLysValGluTrpGluGluLeuLysCysSerLeuArgPheSerLeuAsnGlnGlu 798
QY 948 GAAGAGAAGAGAGAAATCGCGATATATTAAATGAAAAAATAGGGAAGATTTAGGAAGA 1007
Db 799 GluGluLysArgArgAsnAlaAspThrLeuTyrgluLysIleArgGluGlnLeuArgArg 818
QY 1008 ATCGAAGAGCAGCATAGGAAGAGTTAGAGTGAACCAACACTTGACAGGCTCTCAGA 1067
Db 819 LysGluGluGlnTyrgluValGluValLysGlnGlnLeuGluLeuSerLeuGln 838
QY 1068 ATACAAGATTAGAAATGCAAGAGTGTAGAAAGTAAATTTGAATTCAGGTT----- 1115
Db 839 ThrLeuGluMetGluLeuArgThrValLysSerAsnLeuAsnGlnValGlnGluArg 858
QY 1115 ----- 1115
Db 859 AsnAspAlaGlnArgGlnLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyIle 878
QY 1115 ----- 1115
Db 879 LeuThrAsnHisLeuSerLysGlnLysGluLeuGluMetAlaGlnLysLysMetAsnSer 898
QY 1116 -----TCTCACACTCATCAAAATGAAATATTATCTTTACATGAAATTCATGTTGAAA 1169
Db 899 GluAsnSerHisSerHisGluGluLysAspLeuSerHisLysAsnSerMetLeuGln 918
QY 1170 AAGGAATTCGATCGATCAAACTGAAATAGCCACTGAACACCAATCAGGAAAG 1229
Db 919 GluGluIleAlaMetLeuArgLeuGluLeuAspThrIleLysAsnGlnAsnGlnLys 938
QY 1230 GAAATATAATATCTTTGAGACATTAAGATTTTAAAGAAAGAAATGCTGAACTCAGATG 1289
Db 939 GluLysLysCysPheGluAspLeuLysIleValLysGluLysAsnGluAspLeuGlnLys 958
QY 1290 ACCCTAAACTGAAAGAGAAATCATTAATAAGGGCATCTCAATATAGTGGCAGCTT 1349
Db 959 ThrIleLysGlnAsnGluGluThrLeuThrGlnThrIleSerGlnTyrgluLysGluLeu 978
QY 1350 AAGATTCGATAGCTGAGAACCAATGCTCACTTCTTAAATTTGAAG---GAAAAACAAGAC 1406
Db 979 SerValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGluLysGlnSer 998
QY 1407 AAGAAATATCTAGAGCGAAATTAATCAATCACCATCTAGACTGGCTTCTGCTGTACAA 1466

```

```

Db 999 LysGluArgLeuGluAlaGluValGluSerTyHisSerArgLeuAlaAlaIleHis 1018
QY 1467 GACCATGATCAAAATTTGTGACATCAAGAAAAGTCAAGAACCTGCTTTCACATTCGACGA 1526
Db 1019 AspArgGlnSerGlnThrSerLysArgGluLeuGluLeuAlaPheGlnArgAlaArg 1038
QY 1527 GATGCTTGT-----TTGCAAAAGAAAATGAATGTTGATGTGAGTAGTACCCGATATATA 1580
Db 1039 AspGluCysSerArgLeuGlnAspLysMetAsnPheAspValSerAsnLeuLys-AspAs 1058
QY 1581 CAATGAGGTGCTCATCAACCACTTTCTTGAAGCTCAAGAAATCCANAAGCCTTAAAT 1640
Db 1058 nAsnGluIleLeuSerGlnGlnPheLysThrGluSerLysLeuAsnSerLeuGluI 1078
QY 1641 TAATCTCAATTATGACGAGATGCTCTAAGAGAAAATACATGTTTCAGGAACATGCAC 1700
Db 1078 eGluPheHisThrArgAspAlaLeuArgGluLysThrLeu-GlyLeuGluArgValG 1098
QY 1701 AAAGAGACCAACCGTGAACACACAGTGTCAAAATGAAGAACTGAACACATGTTTCAAANC 1760
Db 1098 nLysAspLeuSerGlnThrGlnCysGlnMetLysGluMetGluGlnLysTyrglnAsnG 1118
QY 1761 AACAGATGATGTGAACACACACACTGANCAGCAGGAGTCTCTAGATCAGAAATATTTC 1820
Db 1118 lGlnValLysValAsnLysTyrglyLysGlnGlnSerValGluGluArgLysSerG 1138
QY 1821 AACTACAAAGCAAAATATGTGCTTCAACAGCAATTAGTTCATGACACATAANGAAAGCT 1880
Db 1138 lLeuGlnSerGluAsnMetLeuLeuArgGlnGlnLeuAspAspAlaHis-AsnLysAla 1157
QY 1881 GACACAAAGCAAGATAACATTGAT-----NTTCATTNTCTTGAGAGGAAA 1928
Db 1158 AspAsnLysGluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLys 1177
QY 1929 ATGCNCATCAT-----CTTCTAAAGAGAAAAATGAGAGATATTT 1969
Db 1178 LeuGlnAlaGluSerGluLysGlnSerLeuLeuLeuGluGluArgAsnLysGluLeuIle 1197
QY 1970 NATTCATNATAACCATTTAAACCCGATATATTTCAATATGGAAGAAAAAANAANA 2029
Db 1198 SerGluCysAsnHisLeuLysGlu-ArgGlnTyrglnTyrgluAsnGluLysAlaGlu 1217
QY 2030 A 2030
Db 1217 g 1217
RESULT 6
Q9HCD1 PRELIMINARY; PRT; 718 AA.
ID Q9HCD1
AC Q9HCD1; 2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein KIAA1641 (Fragment).
GN KIAA1641.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
ON NCBI_TaxID=9606;
RX SEQUENCE FROM N.A.
RP MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro."
RL DNA Res. 7:273-281(2000).
DR EMBL; AB046861; BAB13467.1; -.
KW Hypothetical protein.
FT NON TER
SQ SEQUENCE 718 AA; 82972 MW; 2AD05ADE70E3C97F CRC64;
Alignment Scores:

```

Pred. No.:	1-83e-40	Length:	718
Score:	790.50	Matches:	233
Percent Similarity:	53.83%	Conservative:	132
Best Local Similarity:	34.37%	Mismatches:	242
Query Match:	23.23%	Indels:	71
DB:	4	Gaps:	16
US-09-602-362E-15 (1-2030) x Q9HCD1 (1-718)			
QY	111	GAGAGCCATCGCTTCAGAGCTGCCATGAAATGCAAAAGCTGTTCCAAATAAAGCC	170
DB	13	GlnLysGlnProAlaLeuLysAlaThrSerAspLysLysAspValSerAsnLeuPro	32
QY	171	TTGGAAATGAAGATGAACAAACATTCAGAGCAGATGAGATACCTCCCATCAAAATCCAA	230
DB	33	ThrGluLeuLysAspGlyGlnGln-----SerGlyThrValSerSer	46
QY	231	CAAAAGGACTATGAAGAAAGTCTTGGGATTCGAGACTCTCTGTGAGACTGTTTCACAG	290
DB	47	GlnLys-----GlnProAlaThrLysAlaThrSerVal-----Lys	58
QY	291	AAGGATGTGTGTACCCCAAGGCTACACATCAAAAGAAATAGATAAAATAAATCGAAAA	350
DB	59	LysAspSerValSerAsnLeuAlaThrGluLeuLysAsp---GlyGlnLeuArgGlyThr	77
QY	351	TTAGAAGAGTCTCTGATAATGATGTTTTCGAAAGGCTCCCTGCAGAAATGAAGTTCT	410
DB	78	Val-----SerProGlnLysGlnSerAlaGlnLysValIlePheLysLysLysValSer	95
QY	411	ATTCCAACTAAAGCCTTAGAATGTAGGACATGCAACCTTCAAGACGAGCCTCCCGAG	470
DB	96	LeuLeuAsnLeuAlaThrArgIleThrGlyGlyTrpLysSerGlyThrGluTrpProGlu	115
QY	471	AAGCCATCTGCCTTCGAGCTCCATTTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTG	530
DB	116	AsnLeuProThrLeuLysAlaThrIleGluAsnLysAsnSerValLeuAsnThrAlaThr	135
QY	531	GAATTGAAGATGAACAAACATTCAGAGCAGATCAG-----ATGTTCCCTTCAGATCA	584
DB	136	LysMetLysAspValGlnThrSerThrProGluGlnAspLeuGluMetAlaSerGluGly	155
QY	585	AAACAAAAGAGTGAAGAAATCTTGGGATTCGAGAGTCTCCGTCAGACTGTTTCA	644
DB	156	GluGlnLysArgLeuGluGluTrpGluAsnAsnGlnProGlnValLysAsnGlnIleHis	175
QY	645	CAGAAG-----GATGTGTGTACCCAGGCTACACATCAAAAGAA-----	686
DB	176	SerArgAspAspLeuAspAspIleGlnSerSerGlnThrValSerGluAspGlyAsp	195
QY	687	-----ATGGATAAATAAGTGGAAATTA	710
DB	196	SerLeuCysCysAsnLysAsnValIleLeuLeuLeuAspGlnHisGluMetLysCys	215
QY	711	GAAGATTCAACTAGCTATCAAAATCTTGGGATTCAGACTGTTTCAAGAGCAAGG	770
DB	216	LysAspCysValHisLeuLeuLysLysLysAsnThrPheCysLeuTrpLysArgLeuIle	235
QY	771	GAACATCAAAAGATCAGCTGCAACACGTCAGAGAAATGGAACAAATGAAAGAAAG	830
DB	236	LysLeuLysAspAsnHisCysGluGlnLeuArgValLysIleArgLysLeuLysAsnLys	255
QY	831	TTTTGTGTACTGAAAGAAAGTCTCAGAGCAAAAGAAATTAATCACTAGAGAAC	890
DB	256	AlaSerValLeuGlnLysArgIleSerGluLysGluGluLeuLysSerGlnLeuLysHis	275
QY	891	CAAAAGCTTAATGGCAACAGCTCTCCAGTGTGACATTCAGCTTTTAAACACAGAGAA	950
DB	276	GluIleLeuGluLeuGluLysGluLeuCysSerLeuArgPheAlaIleGlnGlnLys	295
QY	951	GAGAGAGAGAAATGCCCATATATTAATGAAATATAGGAGAAATTAGGAGAAATC	1010
DB	296	LysLysArgAsnValGluGluLeuHisGlnLysValArgGluLysLeuArgIleThr	315

QY	1011	GAGAGCAGCATAGGAAAGAGTTAGAAAGTGAAACAACTTGAAACAGGCTCTCAGAATA	1070
DB	316	GluGluGlnTrpArgIleGluAlaAspValThrLysProIleLysProAlaLeuLysSer	335
QY	1071	CAAGATATAGAAATGAAGAGTGTAGAAAGTAATTTGAATCAGGTTTCTCACACTATGAA	1130
DB	336	AlaGluValGluLeuLysThrGlyGlyAsnAsnSerAsnGlnValSerGluThrAspGlu	355
QY	1131	AATGAAATATCTCTTACATGAAATTTGATGTTTCAAAAGGAAATTCCTGCTGCTGCTG	1190
DB	356	LysGluAsp---LeuLeuHisGluAsnArgLeuMetGlnAspGluLeuAlaArgLeuArg	374
QY	1191	CTGGAATAGCCACACTGAAACCAATACCAAGAAAGAAAGAAATTAATCTTTGAGGAC	1250
DB	375	LeuGluLysAspThrIleLysAsnGlnAsnLeuGluLys-----LysTrpLeuLysAsp	392
QY	1251	ATTAAGATTTTAAAGAAAGATGCTGAACTTCAGATGACCTCAAACTGAAAGAGGAA	1310
DB	393	PheGluLeuValLysArgLysHisGluAspLeuGlnLysAlaLeuLysArgAsnGlyGlu	412
QY	1311	TCATTAACTAAAGGCGCATCTCAATATAGTGGGCGCTTAAAGTTCTGATAGCTGAGAAC	1370
DB	413	ThrLeuAlaLysThrIleAlaCysTrpSerGlyGlnLeuAlaAlaLeuThrAspGluAsn	432
QY	1371	ACAATGCTCACTCTTAATG---AAGGAAACAAAGACAAAGAAATATCTAGAGGAGAA	1427
DB	433	ThrThrLeuArgSerLysLeuGluLysGlnArgGluSerArgGlnArgLeuGluThrGlu	452
QY	1428	ATTGAATCACACCATCTAGACTGCGTCTGCTGTACAAGACCATCAAACTTGTCAG	1487
DB	453	MetGlnSerTrpHisCysArgLeuAsnAlaAlaArgCysAspHisAspGlnSerHisSer	472
QY	1488	TCAGAAAAAGTCAAGAACCTGCTTTCCACATTCGAGAGATGCTTGT-----TTGAAA	1541
DB	473	SerLysArgAspGlnGluLeuAlaPheGlnGlyThrValAspLysCysArgHisLeuGln	492
QY	1542	AGAAAAATGAATGTTGATGTGATAGTACCGATATATATAACAATGAGTCTCCATCAAC	1601
DB	493	GluAsnLeuAsnSerHisVal-Leu-----IleLeuSerLeuGlu	505
QY	1602	ACTTCTCGAAGCTCAAGGAAATCCANAAGCCTTAAATAATTAATCTCAATATGCGAGGA	1661
DB	505	nLeuSerLysAlaGluSerLysSerArgValLeuLysThrGluLeuHisTrpThrGlyGlu	525
QY	1662	TGCTCTAAGAGAAATACATGCTTTCAGGACATCCCAAGACAGACCAACGCTGAAAC	1721
DB	525	uAlaLeuLysGluLysAlaLeuValPhe-GluHisValGlnSerGluLeuLysGlnLysG	545
QY	1722	AGTGTCAAATGAAGGAGCTGAACACATGTTNTCAANCGAACCAAGATNATGTCAACAAAC	1781
DB	545	InSerGlnMetLysAspIleGluLysMetTrpLysSerGlyTrpAsnThrMetGluLysC	565
QY	1782	ACACTGACGAGGAGTCTCTAGATCAGAAATATTTCAACTACAAACCAAAATATCT	1841
DB	565	ysIleGluLysGlnGlu-----ArgPheCysGlnLeuLysLysGlnAsnMetL	581
QY	1842	GGCTTCAACAGCAATTAGTTTCATGACATGAAGAAAGCTGACAAACAAAGCAAGTAACA	1901
DB	581	euleuGlnGlnGlnLeuAspAspAlaArg-AsnLysAlaAspAsnGlnGluLysAlaIle	600
QY	1902	ATTGATNTTCATNT-----CTTCAG-----AGGAAA	1928
DB	601	LeuAsnIleGlnAlaArgCysAspAlaArgValGlnAsnLeuGlnAlaGluCysArgLys	620
QY	1929	ATGCNATCATCTCTTCAAAAGAGAAAAATCAGAGATATTTTATACNATAACCAATTA	1988
DB	621	HisArgLeu-LeuLeuGluGluAspAsnLysMetLeuValAsnGluLeuThrHisSerL	640
QY	1989	AAAACCCGTATATTTCAATATGGAATAAAAAAANAAAAA	2030
DB	640	ysGlu-LysGluCysGlnTrpGluLysGluLysAlaGluArg	653

Qy	1944	TAAAGAGAAAAATGACGAGATATTNATTACNATAACCATTTTAAAAAACCCGTATATTT	2003
Dd	399	eUGInGuLySAsnLySGluLeuMetAspGluTyraSnHisLeuLySGlu-ArgMetAsp	418
Qy	2004	CAATATGGAIAAAAAAAAAA	2021
Dd	419	GlnCysGluLySGluLyS	424
 RESULT 9 Q8IZM7 PRELIMINARY; PRT; 453 AA.			
ID	Q8IZM7		
AC	Q8IZM7		
DT	01-MAR-2003	(TrEMBLrel. 23, Created)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last sequence update)	
DT	01-MAR-2003	(TrEMBLrel. 23, Last annotation update)	
DE	Melanoma-associated antigen (Fragment).		
OS	Homo sapiens (Human).		
OC	Eumycota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
CX	NCB_I_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Bruno R., d'Orlando O., Altomonte A., Lama J. E., Maio M., Pucillo C.;		
RT	"Study of the immune profile in metastatic melanoma patients immunized		
RT	with anti-idiotypic antibody by SEREX analysis."		
RL	Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF543495; AAK40505.1; -.		
FT	NON TER		
SQ	SEQUENCE 453 AA; 53140 MW; 81DF80012122BBCA CRC64;		

DR Pfam; PF00023; ank; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 996 AA; 116065 MW; 758060E3D8A1543 CRC64;

Alignment Scores:
Pred. No.: 2,53e-19 Length: 996
Score: 443.50 Matches: 131
Percent Similarity: 50.84% Conservatives: 80
Best Local Similarity: 31.57% Mismatches: 164
Query Match: 13.03% Indels: 40
DB: 4 Gaps: 6

US-09-602-362E-15 (1-2030) x Q8IVF6 (1-996)

QY 815 ACAATGAAAGAAAGTTTGTGCTACTGAAAGAAAGAACTGTGAGA----- 859
Db 239 ThrAlaGluAspTyrAlaLeuGluCysSerAspLeuArgSerIleArgGlnGlnLeuGlu 258
QY 860 -----AGCAAGAAAGAAATCAACAGTTAGAGAACCAAAAGTTAAATGGCAACAGA 913
Db 259 HisLysAsnLysMet-LeuLysAsnHisLeuArgAsnAsp----- 271
QY 914 GCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCCGATAT 973
Db 272 -----AsnGlnGluThrAlaAlaMetLysProAla----- 281
QY 974 ATTAAATGAAAGAAATAGGAAGAAATAGGAAGAAATAGGAAGAGAGAGAGAGAGTT 1033
Db 282 -----AsnLeuLysLysArgLysGluArgAlaLysAlaGluHisAsnLeuLysValAla 300
QY 1034 AGAAGTGAACCAACAACTGACAGGCTCTCAGAAATACAGATATAGAAATGGAAGTGT 1093
Db 300 rGluGluLysGlnGlu-----ArgLeuGlnArgSer 310
QY 1094 AGAAGTAAATTAATGATCAGTTTCTCAGTCACTCATGAAATGAAATATCTCTTACATGA 1153
Db 310 rGluAsnLysGlnProGlnAspSerGlnSerTyrGlyLysLysLysAspAlaMetTyrG 330
QY 1154 AAATTGCTATTTGAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1213
Db 330 yAsnPheMetLeuLysLysAspIleAlaMetLeuLysGluGluLysTyrAlaIleLysAs 350
QY 1214 CCAATACCGAGAAAGAAATTAATTAATTTGAGGACATTAAGATTTTAAAGAAAGAA 1273
Db 350 nAspSerLeuArgLysGluLysLysTyrIleGlnGluIleLysSerIleThrGluIleAs 370
QY 1274 TGCTGAATTCAGATGACCTCAAACTGAAAGAGGAATCATTAACATAAGAGGCGATCTCA 1333
Db 370 nAlaAsnPheGluLysSerValArgLeuAsnGluLysMetIleThrLysThrValAlaAr 390
QY 1334 ATATAGTGGGACCTTAAAGTTCTGATAGTGAAGACACAACTGCTCACTTCTAAATG-- 1391
Db 390 gTyrSerGlnGlnLeuAsnAspLeuLysAlaGluAsnAlaArgLeuAsnSerGluLeuCl 410
QY 1392 -AAGGAAACCAAGCAAGAAATATCTAGAGGAGAGAAATGAAATACACACATCTAGACT 1450
Db 410 uLysGluLysHisAsnLysGluArgLeuGluAlaGluValGluSerLeuHisSerSerLe 430
QY 1451 GGCTTCTGCTGTCAAGACCATCATCAATTTGTGACATCAAGAAAGTCAAGAACCTGCTG 1510
Db 430 uAlaThrAlaIleAsnGluLysAsnGluIleVal--GluArgLysAspLeuGluLeuVa 449
QY 1511 TTTCACATTCGAGGAGATCTTGTTCGAAAGAAATGAATGTTGATGTGAGTAGTAC 1570
Db 449 lLeuTrpArgAlaAspValSerArgHisGluLysMetGlySerAsnIleSerGlnLe 469
QY 1571 CGATATATAACAATGAGTGTCTCCATCAACCACTTTCTGAAGCTCAAGAAATCCANAA 1630
Db 469 u-ThrAspLysAsnGluLeuLeuThrGluGlnValHisLysAlaArgValLysPheAsnT 489

QY 1631 GCCTAAAAATTAATCTCAATTATCGAGGATGCTCTTAAGAGAAATATACATTGTTTCAG 1690
Db 489 hrLeuLysGlyLysLeuArgGluThrArgAspAlaLeuArgGluLysThrLeuAlaLeuG 509
QY 1691 GAACATGCACAAAGAGACCACGCTGAACACAGTGTCAATGAGGAGAGTGAACACATG 1750
Db 509 lyser-ValGlnLeuAspLeuArgGlnAlaGlnHisArgIleLysGluMetLysGlnMet 528
QY 1751 TNCAAANCGAACCAAGATNATGTCAACAAACACACACTGACGAGGAGTCTCTAGATCAG 1810
Db 529 HisProAsnGlyGluAlaLysGluSerGlnSerIleLysGlnAsnSerLeuGluGlu 548
QY 1811 AATATTATCACTACAAACAAATATGCTGCTTCAACAGCAATAGTTCATGTCACAT 1870
Db 549 ArgIleArgGlnGlnGluLeuGluAsnLeuLeuGluArgGlnLeuGluAspAlaArg 568
QY 1871 AANGAAGCTGACAAACAAAGCAAGATTAACAATTTGATTTTCAATTCTTTCAGAGGAAAT 1930
Db 569 LysGluLysAspAsnLysGluIleValIleAsnIleHisArgAspCysLeuGluAsnGly 588
QY 1931 GCNCATCATCTTCTAAAGAGAAATAGGAGATATTTTATACNATACCATTTAAATA 1990
Db 589 LysGluAspLeuLeuGluGluArgAsnLysGluLeuMetLysGluTyrAsnTyrLeuLys 608
QY 1991 AACCCGTATATTCAATATCGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2029
Db 609 GluLysLeuLeuGlnCysGluLysGluLysAlaGluArg 621

RESULT 11
Q8T5C7 PRELIMINARY; PRT; 2055 AA.
AC Q8T5C7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Chimeric erythrocyte-binding protein MAEBL.
GN MAEBL.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Blair B.L., Kappe S.H.T., Maciel J.B., Balu D.J., Adams B.;
RT "Plasmodium falciparum mael is a unique member of the ebl family."
RL Mol. Biochem. Parasitol. 0:0-0(2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Michon P., Stevens J.R., Kaneko O., Adams J.H.;
RT "Plasmodium falciparum mael is a unique member of the EBL family."
RL Mol. Biol. Evol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Ghai M., Dutta S., Ockenhouse C.F.;
RT "Identification, expression and functional characterization of MAEBL,
RT a sporozoite and asexual blood stage chimeric erythrocyte binding
RT protein of Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 0:0-0(2002).
DR EMBL; AY042084; AAL10509.1; -
DR EMBL; AF400002; AAM90625.1; -
DR HSP; P04268; IIC2.
DR InterPro; IPR008602; Duffy binding.
DR Pfam; PF05424; Duffy-binding; 1.
SQ SEQUENCE 2055 AA; 243210 MW; 783F48BB7286B1F3 CRC64;

Alignment Scores:
Pred. No.: 1.98e-12 Length: 2055
Score: 330.00 Matches: 171
Percent Similarity: 42.69% Conservatives: 124
Best Local Similarity: 24.75% Mismatches: 260
Query Match: 9.70% Indels: 136
DB: 5 Gaps: 27

US-09-602-362E-15 (1-2030) x Q8T5C7 (1-2055)

US-09-602-362B-15 (1-2030) X Q815C7 (1-2055)

Qy	15	GATGGTCTCTGAAGGCTAACTCGGAATGAAAGTTTCTATTCCAATAAGCCTTAGAA	74
Db	1264	AspGluLeuLysLysAlaGluGluLysLysAlaAspGluLeuLysSerGluGlu	1283
Qy	75	TTGATGACATGCAACTTTCAAGCAGAGCCTCCGAGAGCCATCTCCCTCGAGCCT	134
Db	1284	LysLysAlaAspGluLeuLysLysAlaGluGluLysLysAlaAspGluLeu	1302
Qy	135	GCAATTGAAATGCAAAAGTCTGTTCAAAATAAGACCTTGGAAATGAAGAT	188
Db	1303	--LeuLysLysAlaGluGluLysLysAlaAspGluLeuLysLysAlaGlu	1321
Qy	189	CAACATTGAGACGAGTGAATACTCCCATCAGAATCCAACAAGAGCTATGAAGA	248
Db	1322	GluLysLysAlaAspGluValLysLysAlaGluGluLysLysAlaAspGluLeu	1341
Qy	249	AGTTCTTGGAGTCTCTGAGACTCTCTGAGACTGTTTCACAGAAGGATGTGT	308
Db	1342	LysLysSerGluGluLysLysAlaAspGluLeuLysLysLysAlaAsp	1354
Qy	309	AGGCTACACATCAAAAGAAATAGATAAATAATGGAATAATAGAGAGTCTCTGAT	368
Db	1355	LysSerGluGluLysLysAlaAspGluLeuLysLysAlaGluGluLysLys	1374
Qy	369	AATGATGTTTCTGAAGGCTCCTGAGAGTCTGAGACTTTCATTCCAATAAGCCTTA	428
Db	1375	AlaAspGluLeuLysLysLysLysLysLysLysLysLysLysAlaAsp	1389
Qy	429	GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAG	488
Db	1390	GluLeuLysLysLysLysLysLysLysLysLysLysAlaAspGluLeuLys	1405
Qy	489	CCTGCCATTGAAATGCAAAAGTCTGTTCCAAAATAAGCCTTGAATTTGAAGAT	542
Db	1406	LysLysAlaGluGluLysLysLysLysLysLysLysLysLysLysAla	1421
Qy	543	GAACAACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAGAGCTTGA	602
Db	1422	GluGluLysLysLysAlaAspGluLeuLysLysLysLysLysLysLysAlaGlu	1435
Qy	603	GAAATTTCTTGGGATCTCGAGAGTCTCCGTGAGACTGTTTCACAGAAG	656
Db	1436	GluLysLysLysLysAlaGluAsnLeuLysLysAlaGluGluLysLysAlaAspGlu	1454
Qy	657	TGTTGATCCCAAGGCTACACATCAAAAAGAAATGATATAATAGTGAAATTAGAAGAT	716
Db	1455	LeuLysLysLysAlaGluGluLysLysAlaAspGluLeuLysLysLysAlaGluGlu	1474
Qy	717	TCAACTAGCCTATCAAAAATCTTGGATACAGTTTCATTCTTGTGAAAGACAAAGGA	776
Db	1475	LysLysLysAlaAspGluLeuLysLysAlaGluGluLysLysLysAlaAspGluLeu	1494
Qy	777	CAAAAAGATCACTGTGACACAGCTACAGGAAAATGGAA	830
Db	1495	LysLysAlaGluGluLysLysLysAlaAspGluLeuLysLysAlaGluGluLysLys	1514
Qy	831	TTTTGTGTACTGAAAAG	866
Db	1515	AlaAspGluLeuLysLysAlaGluGluLysLysLysAlaAspGluLeuLysLysAlaGlu	1534
Qy	867	GAAATAAATCAGCTTAGAACAACCAAAAGTT	917
Db	1535	GluLeuLysLysAlaGluGluLysLysValGluGlnLysLysArgGluGluArg	1554
Qy	918	TGCAGTGTGAGATTG	944
Db	1555	ArgAsnMetAlaLeuArgAlaGluLeuLysGlnLeuGluLysLysArgIleGlu	1574
Qy	945	-----GAAGAGAGAGAGAGAAATCCGATATATTAAATGAAAA	986
Db	1575	GluValMetLysLeuTyGluGluGluLysMetLysAlaGluGlnLeuLysLysGlu	1594

RESULT 12

Q81HP3 TD 09TH

Q81H	AC
Q81H	AC

PRELIMINARY:

4
3
2
1
4
3
2
1

QY 1527 GATGCTTGTTCGAAAGAAATGAATGTTGATGTGAGTAGTACCGATATATACAATGA 1586
 Db 1747 ----- 1747
 QY 1587 GGTGCTCCATCAACACATTTCTGAGCTCAAGGAATCCANAAGCCTAAATAATTAATCT 1646
 Db 1748 -----LysalaGluGluIleAa-GLysGluLysGlu--Alav 1759
 QY 1647 CAATTATGAGGAGATGCTCTAAGAGAAATACATTTGTTTCAGGAACATCACAAAGAG 1706
 Db 1759 alileGluGluLys-----LysLysGluaspGluLysA 1771
 QY 1707 ACCACGTAACACAGAGTGTCAATGAAGGAGCTGAACACATGNTCAAACGACAAAG 1766
 Db 1771 tGArgMetGluValGluLysIleLysAspThrLysAspAsnPheGluAsnIleGlnG 1791
 QY 1767 ATNATGTGAACAAACACACT-----GANCACGAGAGTCTCTAGATCAGAAATTAATTC 1820
 Db 1791 luGluAsnAsnLysAsnThrProTyrIleAsnLysGluMetPheAspSerGluLysG 1811
 QY 1821 AACTCAAA---AGCAAAATATGTGGCTTCAACAGCAATTAATTCATGCACATANGAAA 1877
 Db 1811 luValIleThrLysAsnMetGlnLeuAsnGluAlaAspAlaPheGluLysHisAsnS 1831
 QY 1878 GCTGACAAACAAAGCAAGATAACAATGATNTTCATTCTTGAGAGGAAATGCNCATC 1937
 Db 1831 erGluAsnSerLysSerSerAsnLysAsnAlaAspPheSerLysGluLysasp----- 1848
 QY 1938 ATCTCTTAAAGAGAAATAGGAGATA 1966
 Db 1849 --LeuLeuGluaspIleGluAsnIle 1857
 RESULT 13
 O76329 PRELIMINARY; PRT; 1738 AA.
 AC O76329;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Interactin.
 GN ABPD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 ON NCBI_TaxID=44689;
 RN [1]_
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98365468; PubMed=9700162;
 RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;
 RT "Interactin, an actin-binding protein of the alpha-actinin superfamily
 RT in Dictyostelium discoideum, is developmentally and cAMP-regulated and
 RT associates with intracellular membrane compartments.";
 RL J. Cell Biol. 142:735-750(1998).
 DR EMBL, AF057019; AAC34582.1; --
 DR FIR; T14867; T14867.
 DR HSP; P46939; IQAG.
 DR GO; GO:0003779; F:actin binding; IEA.
 DR InterPro; IPR001589; Actbind actinin.
 DR InterPro; IPR001715; Calponin-like.
 DR InterPro; IPR001990; Granin.
 DR InterPro; IPR001451; Hexapep_transf.
 DR InterPro; IPR002017; Spectrin.
 DR Pfam; PF00307; CH; 2.
 DR SMART; SM00033; CH; 2.
 DR PROSITE; PS00019; ACTININ 1; 1.
 DR PROSITE; PS00020; ACTININ 2; 1.
 DR PROSITE; PS50021; CH; 2.
 DR PROSITE; PS00422; GRANINS 1; 1.
 DR PROSITE; PS00101; HEXAPEP-TRANSFERASES; 1.
 SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Alignment Scores:
 Pred. No.: 1.01e-11 Length: 1738

Score: 318.50 Matches: 180
 Percent Similarity: 37.86% Conservative: 146
 Best Local Similarity: 20.91% Mismatches: 244
 Query Match: 9.36% Indels: 291
 DB: 5 Gaps: 33
 US-09-602-362E-15 (1-2030) x O76329 (1-1738)
 QY 69 TTGAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAGCCATCTGCCTTC 128
 Db 585 LeuGlnLeuGlnAspGlnGlnAspIleLys-----GluLysGluPheGlnPhe 600
 QY 129 GAGCCT-----GCCATTGAAATGCAA 149
 Db 501 GluLysGlnGlnLeuLeuSerGlnIleAspSerIleThrThrAsnIleGlnGluTyrGln 620
 QY 150 AAGTCTGTTCCAAATAAAGCCTTGGAAATGAAGAATGAACAACATTCAGAGCAGATGAG 209
 Db 621 AspLysPheAsnAsnLeuGlnGlnGluPheAsnThrGlnGlnThrLeuAsnGlnGlnGlu 640
 QY 210 ATA-----CTCCCATCAGAAATCCAAACAA-----AAGGACTATCAAGAAAGTTCTTGG 257
 Db 641 ThrHisArgLeuThrGlnGlnLeuTyrGlnIleAsnThrAspTyrAsnGluLysGlnThr 660
 QY 258 GAT-----TCTGAGAGTCTCTGTGAGACTGTT 284
 Db 661 GlnLeuGlnSerGluIleLysAspAsnGlnThrIleAsnGlnGlnLeuAsnLysGlnLeu 680
 QY 285 TCACAGAGAGTGTGTTTACCCAGGCTACACATCAAAAGAA-----ATAGAT 335
 Db 681 SerGluLysAspLysGluIleGluLysLeuSerAsnGlnGlnGlnGlnGlnGlnGlnGlu 700
 QY 336 AAAATAAATGGAATAATAGAAAGAGTCTCTGATGATGATGATGATGATGATGATGATGAT 395
 Db 701 LysIleAsnAsnLeuLeuLeuLysGluLysAspCysLeuIleGlu----- 717
 QY 396 AGAATGAAGTTCTTATTCACAACTAAAGCCTTAGAA----- 431
 Db 718 -----ArgIleAsnGlnGlnLeuLeuLeuAsnIleAspLeuAsnSerLysTyr 733
 QY 432 -----TTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAGCCATCTGCC 482
 Db 734 GlnGlnLeuLeuLeuGluPheGluAsnPhelys----- 744
 QY 483 TTCAGCCTGCGCATGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGAATTCAGAAAT 542
 Db 745 -----LeuAsnSerSerLysGluLysGluAsnGlnLeuAsnGlnLeuGlnSer 760
 QY 543 GAACAAACATTCAGAGCAGATCAGATCTTCCT----- 575
 Db 761 LysGlnAspGluArgPheAsnGlnLeuAsnAspGluLysLeuGluLysGlnLeu 780
 QY 576 -----TCAGAAATCAAAACAAAGAGGTGGAAGAAATCTTGG 614
 Db 781 GlnSerIleGluAspGluPheAsnGlnTyrLysGlnGlnGlnLeuSerSerAsnSerAsn 800
 QY 615 GATTCAGAGTCTCCGTGAGACTGTTTCACAGAGAGATGTGTGTGTACCCAGGCTACA 674
 Db 801 IleAspGlnGlnLeuGlnSerThrIleIleGlu-----LeuSerGluLeuLys 816
 QY 675 CATCAAAAGAAATGAGT-----AAAATAAGTGGAAAATTAGAAGATTCAATAGCCTA 728
 Db 817 GluGlnLysGluLeuAsnAspSerLysLeuIleGluLysGluLysGlnLeuGlnGlnLeu 836
 QY 729 TCAAAATCTTGGATACAGTTTCATCTTGTGAAGACCAAGGAGACTTCAAAAGATCAC 788
 Db 837 GlnGlnGluPheAspGlnLeuAsn-----GluLysAsnGlnLysAspHis 851
 QY 789 TGTGAACAACGTACAGGAAATAATGCAAAATGAAAGAGATTTTGTGTACTGAAAGAG 848
 Db 852 GlnAspGln-----LeuGluLeuLeuGluLysGlnLeuLysGlnLeuGlnGln 867
 QY 849 AAACCTGTGAGAGCAAGAAATAAATACAGTTAGAACCAAAAGTTAAATGG--- 905

```
Db      868  GluTyrAspGlnLeuAsnGluThrAsnGlnSerIleGluAsnGlnLeuAsnGlnGlnAsn 887
      906  -----GAACAGAGCTCTGCAGTGTGAGATTGACT 935
      888  LeuIleAsnLysGluAsnLeuAsnGluLysGluGlnGlnLeuLysGluGlnAsnGln 907
      936  TTAACCAAGAAGAAGAGAGAAGA-----962
      908  LeuAsnGlnGlnIleGluLysIleGlnPheAspGlnGlnPheSerLysGlnAsnSer 927
      963  --AATCCCATATATAATGAATA-----ATTAGGAAGAAATTA 1001
      928  IleAsnIleGluLeuValAsnGluLysAsnGluLysLeuIleGlnLeuGlnGlnAspTyr 947
      1002  GGAAGAAATCAAGAGACGATAGG-----1025
      948  AspGlnLeuLysGlnGlnAsnArgSerAsnAspGluLysAspGluAsnAspLeuIleGlu 967
      1025  -----1025
      968  LysGluAsnGlnLeuLysSerIleGlnAsnGlnLeuAsnGlnLeuIleGluLysAsnGlu 987
      1026  -----AAAGAGTTAGAGTGAACAACAA-----CTTGAACAGGCTCTCAGAAATACAA 1073
      988  SerAspHisLysGluGlnGlnLeuLysGlnGlnSerIleGluAsnAspLeuIleGluLys 1007
      1074  GATATAGAATTGAAGAGTGTAGAAAGTAATTG-----AAT 1109
      1008  GluAsnGlnIleGlnLeuGlnSerGlnLeuAsnGluGlnArgGlnGlnGlnGlnSerAsn 1027
      1110  CAGGTTTCACACTCATGAAATGAAATATCTCTTACAT-----1151
      1028  GlnLeuSerGluLysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGlnPheAspGln 1047
      1152  -----GAAATTTGCTGTTGAAAGAAATTTGCC 1181
      1048  LysGluGlnGlnLeuLysGlnGlnSerIleGluAsnAspLeuPheGluLysGluAsnGln 1067
      1182  ATCTTAAGTGAATAGCAGTGAACACCAATACAGAGAAAGAAATTAATATAC 1241
      1068  IleGlnGlnLeuGln---SerGlnLeuAsnGluGlnArgGlnGlnSerAsnGlnLeu 1086
      1242  TTTGAG-----GACATTAAAGATTAAAGAAAGAAATGCTGAACTTCAGATGACC 1292
      1087  SerGluLysAspGlnGlnLeuAsnGlnLeuIleGluLysAsnGluSerAspGlnLysGlu 1106
      1293  CTAAGACTGAAGAGGAATCA-----TTAAGTAAAGGCGATCTCAATATAGT 1340
      1107  GlnGlnLeuLysGlnGlnSerIleGluAsnAspLeuIleGluLysGluAsnGlnIleGln 1126
      1341  GGCAGCTTAAAGTTCTGATAGTGAAGACCAATCTCCTCACTTCTAAATTTG-----1391
      1127  ---GlnLeuGlnLeuAsnGlnGluGlnArgGlnGlnSerGluValSerIleAsp 1145
      1391  -----1391
      1146  AsnAspLysIleLeuGluLeuGluLysGlnLeuLysGlnCysGlnSerAspLeuLysLys 1165
      1392  -----AAGAAAACAGACAAAGAAATACTAGAGCGCAAAATGAA-----1433
      1166  LeuAsnAspGluLysGlnGlnAspLysGlnLeuGlnAspLysGlnIleGluPheAsp 1185
      1433  -----1433
      1186  GlnLeuGlnLeuThrPheAsnGlnPheLysAsnAspLysAspSerGlnPheIleGlnLeu 1205
      1434  -----TCACACCATCTAGATGGCTTCTGCTGACAGACCATG-----ATCAATTT 1481
      1206  GlnAspAspGlnLysGlnGlnLeuGlnSerIleGlnGlnAspLeuAsnGlnLeuLysGln 1225
      1482  GTGACATCAAGAAAGTCAAGAACCTG-----CTTTCACATTGCA 1523
```

```
Db      1225  nGluAsnGlnGluLysGluLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGln 1245
      1524  GGAGATGCTTCTTTTGCACAAAGAAAATGAATGTTGATGTAGTAGTATACCGATATATACAA 1583
      1245  n-----PheGluAs 1248
      1584  TGAGGTGCTCCATCAACACATTTCTGAAGCTCAAGGAAATCCANAAGCTTAAATAATTA 1643
      1248  nGlnGluLysGlnLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAs 1268
      1644  TCTCAATATGACGAGATGCT---CTAAGAGAAATACATTTGTTTCAGAA-----1693
      1268  nLeuAsnGlnLeuAsnAspGluAsnGlnGlnGlnLysValLysGlnPheSerGluLysAspG 1288
      1694  -----CATGCACAAAGAGACCAACGTTGAAACAC 1721
      1288  lLysLeuGlnSerIleGlnGlnAspLeuAsnGlnLeuLysGlnGluAsnGlnGluLysG 1308
      1722  AGTGCTCAATGAAGAGAGCTGAACACATGTTTCAAAACGAAAGATNATGTGAACAAAC 1781
      1308  lLysGlnLeuSerGluLysAspGluLysLeuGlnSerIleGlnGlnAspLeuAsnGlnL 1328
      1782  ACATGTCANCAAGCAGGAGCTCTAGATCAGAAATTTTCAACTACAAAGCAAAATATGT 1841
      1328  euAsnAspAspGlnIleLysLysAsnGluLysGluLysGluLysGluGlnLeuL 1348
      1842  GGCTTCAACAGCAATTAAGTTTCATGCATTAAGAAAGCTGCAACAAAGCAAGACAGATA 1901
      1348  ysLeuGlnGlnAspPheAsnAspGlnGlnSerGlnGlnLeuLysGlnLeuGluGlu---1366
      1902  ATTGATTTTCATTTCTTGAGAGGAAATGCNCATCATCTTTTAAAGAGAAATATGAG 1960
      1367  -----LysLeuSerGluLysGluAsnGlnLeuGlnGlnLysGlnGluAsnGlu 1383

RESULT 14
Q80W27  PRELIMINARY; PRT; 520 AA.
AC Q80W27;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testicle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RP "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testicle;
```

RA Strausberg R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC050811; AAH50811.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 520 AA, 61439 MW, 191PF1C27B65ABE42 CRC64;

Alignment Scores:

Pred. No.: 2,09e-11 Length: 520
 Score: 314.00 Matches: 115
 Percent Similarity: 46.04% Conservative: 94
 Best Local Similarity: 25.33% Mismatches: 147
 Query Match: 9.23% Indels: 98
 DB: 11 Gaps: 13

US-09-602-362E-15 (1-2030) x Q80W27 (1-520)

```

QY 603 GAAATCTCTGGAGTCTCGAGAGTCTCCGTGAGACTGTTTCACAG-----AAGGATGTG 656
DB 2 GluAspValGlyAspLeuThrGlnSerSerGluThrAlaSerGluTyrGlyArgProPro 21
QY 657 TGTGTACCCAGGCTACACATCAAAAGAAATGGATAAATAGTGGAAATTAGAAGAT 716
DB 22 CysProAsnTyrGluAlaileLeuArgLeuileGluArgLeuSerLeuGluCysLysAsp 41
QY 717 TCACTAGCTATCAAAATCTTGATACAGTTCATCTTGTGAAGAGCAAGGAACTT 776
DB 42 SerGlySerLeuLeuLysileGlnAsnAlaileHisSerTyrLysArgLeuileGlnLeu 61
QY 777 CAAAAGATCACTGTGAACCACTACAGGAAATATGGAACAAATGAAAGAAAGTTTGT 836
DB 62 LysThrGlyHisCysLysValLeuLeuGluLysLeuGlnThrValGluAsnGluValAsn 81
QY 837 GTACTGAAAGAAACTGTCAAGCAACAAAGAAATAAATACACATTAGAGAACCAAAA 896
DB 82 GlyLeuArgLysLeuAlaAspAla-----AlaArgGluGln 94
QY 897 GTTAAATGGCAACAGAGCTCGCAGTGTGAGATTGACTTTAAACCAAGCAAGAGAGAG 956
DB 95 LeuGlnGlnGlyGlnGluLeuCysAsnValArgPheThrLeuLysGlnGluGluLys 114
QY 957 AGAAGAAATCGCGATATATTAATGAATAAATAGGAAGAAATAGGAAGAAATCGAAGAG 1016
DB 115 ArgLysAsnValValTrpValTyrAspLysMetLysGluGlnLeuArgLysGluAsp 134
QY 1017 CAGCATGAGGAAGTGAAGTGAACCAACCACTGGAACGGCTCTCAGATACACAGAT 1076
DB 135 GlnTyrAsnLysGluValLysMetLysGlnLysLeuGluileArgValArgGluLeuAsp 154
QY 1077 ATAGAATTGAAGAGTGTAGAAAGTAAATTTGAATCAGGTTTCTCACACTCATGAAATGAA 1136
DB 155 MetAspLeuLysThrValArgHisAsnLeuAsnThrVal----- 167
QY 1137 AATTATCTCTTACATGAATTCATGTTGTTGAAAGAAATTTGCCATGCTAAACCTGAA 1196
DB 168 -----LeuGluGluArgAsnCysileGluArgGln----- 177
QY 1197 ATAGCCACACTGAAACCAACTACCAAGGAAAGAAATATAATCTTTGAGGACATTAAG 1256
DB 178 -----LeuPheGlnGluGlnAsnThrArgLleileGlnAsp---Glu 190
QY 1257 ATTTTAAAGAAAGAAATCTGACTTCACTGACCCCTAAACTGAAGAGAAATCATTA 1316
DB 191 IleLeu---AlaAsnHisLeuHisLysGlnLysGluLeuLysMetAlaGlnGluLysile 209
QY 1317 ACTAAAGGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACCAATG 1376
DB 210 -----SerSerGlnLeuGlnGlnAlaProAspGlnHisThrGln 222
QY 1377 CTCACITTTAAATTTGAAGAAACCAAGCAAGAAATACTAGAGCAGAAATTAATCA 1436
DB 223 CysThrGluLysMetArgAspCysileGlnLysLeuGlyLeuGlu----- 237
QY 1437 CACCATCTTAGCTGCTTCTGTGACAGACCATGATCAAAATCTGTGACATCAAGAAAA 1496

```

```

DB 238 -----AsnHisLysLeuLysGluThrAlaArgLys 247
QY 1497 AGTCAAGAACCTGCTTCCACATTGAGGAGATGCTGTTTGGCAAGAAAAATGAATGTT 1556
DB 248 GluAlaGluLys-IleGluGlnLeuLysAsnLeu----- 260
QY 1557 GATGTGAGTAGTACCGATATATTAACAAATAGAGTGCTCCATCAACCACTTCTCAAGCTCA 1616
DB 261 -----SerGlyValLeuSerAspLeuThrGluLysLe 272
QY 1617 AAGGAAATCCANAAGCCTAAAAATTAATCTCAATTATGCGAGGAGATGCTCTAAGAGAAAA 1676
DB 272 uGluSerAlaSerAlaLysTyrLeuHisLeuTyrGluGlnAsnGlnLeuLeuArgG1 292
QY 1677 TACATTGGTTTCAGGACATGCAAGAGACCAACGTAAGAACACAGTGTCAATGAAGG 1736
DB 292 uLeuLeuSerMetArgThr-IleGlnArg-----LysCysGlu---LysL 306
QY 1737 AAGCTGAACATGCTNTCAAANGCAACAGATNATGTGAC----- 1777
DB 306 euLysLysHisLysLysLeuGluGlnGluValSerLeuArgSerHisMetLysA 326
QY 1778 -----AAACACTGANCACAGGAGTCTC 1802
DB 326 xGAsnMetIleGluCysGlyGlnIleGluGlnTyrLysTyrGluileGluGluLysThrL 346
QY 1803 TAGATCAGAAATATTTCACACTACAAAGCAAAATATATGCGCTTCACACAGCAATTAGTTC 1862
DB 346 ysGlnGluLeuValGlyLysLeuLysGlnValAsnLeuPheLeuGlnThrGlnAlaAla 366
QY 1863 ATGCA---CATAANGAAAGCTGACAAACAAAGCAAGAT 1897
DB 366 yrGluAspLysLeuGluLysLeuArgGlnLysGlnAsn 378
RESULT 15
Q15694 PRELIMINARY; PRT; 381 AA.
ID Q15694;
AC Q15694;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Protein immuno-reactive with anti-PTH polyclonal antibodies (Fragment).
DE Homo sapiens (Human).
DE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
DE Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
DE NCBTaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96187433; PubMed=8608414;
RA Kumar R., Haugen J.D., Wieben E.D., Londowski J.M., Cai Q.;
RT "Inhibitors of renal epithelial phosphate transport in tumor-induced osteomalacia and uremia."
RT Proc. Assoc. Am. Physicians 107:296-305(1995).
RL EMBL; U28831; AAB02177.1; --
DR GO; GO:0016021; C:integral to membrane; NAS.
DR GO; GO:0008200; F:ion channel inhibitor activity; NAS.
DR GO; GO:0030319; P:di-, tri-valent inorganic anion homeostasis; NAS.
FT NON_TER 1
FT NON_TER 381
SQ SEQUENCE 381 AA; 42574 MW; BPE144AB3A21C6DD CRC64;
Alignment Scores:
Pred. No.: 1.08e-10 Length: 381
Score: 302.50 Matches: 99
Percent Similarity: 44.29% Conservative: 60
Best Local Similarity: 27.58% Mismatches: 153
Query Match: 8.89% Indels: 47
DB: 4 Gaps: 7
US-09-602-362E-15 (1-2030) x Q15694 (1-381)

```


Search completed: July 15, 2004, 09:24:01
Job time : 142.318 secs

```
QY 12 AAAGATGGTCTTCTGAAGGCTAACTCGGGAATGAAAGTTTCTATTCCAATCAAGCCTTA 71
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 LysGlnProAlaLeuIysAlaThrSerAspLysGluAsnSerValProAsnMetAlaThr 47
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 GAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCCGAGAGCCATCGCTTCGAG 131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 GluThrLysAspGluGlnIleSerGlyThrValSerSerGlnLysGlnProAlaLeuIys 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 CTGCCATTGAAATGCAAAAGTCTGTTCAAATAAAGCCCTTGAATTTGAAGATGAACAA 191
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AlaThrSerAspLysLysAspSerValSerAsnIleProThrGluIleLysAspGlyGln 87
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 ACATTGAGACCATGAGATACCTCCCATCAGATCCAAACAAAGACACTATGAAGNAAGT 251
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 -----GlnSerGlyThrValSerSerGlnLysGlnProAlaTrpLysAlaThr 103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 TCTTGGGATTCAGAGTCTCTGTGAG-----TCTCCT 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 104 SerValLysLysAspSerValSerAsnIleAlaThrGluIleLysAspGlyGlnIleArg 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 279 ---ACTGTTTCACAGAGATGTGTTTACCCAGGCTACACATCAAAAAGAAATAGAT 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 GlyThrValSerSerGlnArgGlnProAlaLeuLysAlaThrGlyAsp---GluLysAsp 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 336 AAAATAAATGCAAAATTAGAAGAG-----TCTCCT 365
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 SerValSerAsnIleAlaArgGluIleLysAspGlyLysSerGlyThrValSerPro 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 366 GATAATGATGTTTCTGAAGGCTCCCTGCAGATGAAAGTTTCTATTCCAATCAAGCC 425
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 163 GlnLysGlnSerAlaGlnLysValIlePheLysLysLysValSerLeuLeuAsnIleAla 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 426 TTAGAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCCGAGAGCCATCGCTTC 485
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 ThrArgIleThrGlyGlyTrpLysSerGlyThrGluTrpProGluAsnLeuProThrLeu 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 486 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTTGAAGATGAA 545
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 LysAlaThrIleGluAsnLysAsnSerValLeuAsnThrAlaThrLysMetLysAspVal 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 546 CAAACATTGAGCACATCAG-----ATGTTCCCTTCAGATCAAAACAAAGAGTT 599
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 223 GlnThrSerThrProGluGlnAspLeuGluMetAlaSerGluGlyGlnLysArgLeu 242
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 600 GAAGAAAATTCCTGGGATTCAGAGTCTCCGTGAGACTGTTTTCACAGAG----- 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 GluGluTrpGluAsnAsnGlnProGlnValLysAsnGlnIleHisSerArgAspAspLeu 262
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 651 ---GATGTGTGTATCCCAAGGCTACACATCAAAAGAA----- 686
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 AspAspIleIleGlnSerSerGlnThrValSerGluAspGlyAspSerLeuCysAsn 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 687 -----ATGGATAAAATAAGTGGAAATTTAGAGATTCAACTAGC 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 CysLysAsnValIleLeuLeuIleAspGlnHisGluMetLysCysLysAspCysValHis 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 726 CTATCAAAATCTTGATACAGTTCACTTCTGTGAAGAGCAAGGGAATTCAAAAGAT 785
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 LeuLeuLysIleLysThrPheCysLeuCysLysArgLeuThrGluLeuLysAspAsn 322
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 786 CACTGTGAACAGTACAGGAAATAAGCAAAATGMAAANGAGTTTGTGTACTGAAA 845
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 HisCysGluGlnLeuArgValLysIleArgLysLeuLysAsnLysAlaSerValLeuGln 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 846 AAGAACTGTGAGACCAAGAAATAATAATCACAGTTAGAGAACCAAAAAGTTAAATGG 905
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 LysArgSerGluLysGluGluIleLysSerGlnLeuLysHisGluThrLeuGluLeu 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 906 GACAAAGACTCTGCAAGTGTGAGATTGACTTTAAACCRAGAGAGAGAGAGAAGA 962
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 GluLysGluLeuCysSerLeuArgPheAlaIleGlnGlnGluLysLysLys 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:45:37 ; Search time 17.1613 Seconds

(without alignments)
12318.669 Million cell updates/sec

Title: US-09-602-362E-15

Perfect score: 3403
Sequence: 1 ctcctgcgttaagatgggt.....gaaaaaaaaaaaaaaaaa 2030

Scoring table:

	BLOSUM62
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool_p/US09602362/runat.15072004.093625.21988/app_query.fasta_1.10325
-DB=SwissProt_42 -QFMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=plo -NORM=ext -HRAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @CGN.1.1.134 @runat.15072004.093625.21988 -NCPU=6 -ICPU=3
-NO_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPF=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	443.5	13.0	992	AN18 HUMAN	Q81v16 homo sapien
2	280.5	8.2	1875	MLP1 YEAST	Q02455 saccharomyc
3	280	8.2	2116	MY52 DICDI	P08799 dictyosteli
4	279	8.2	2017	MY5N DROME	Q99323 drosophila
5	278.5	8.2	3911	AKA9 HUMAN	Q99996 h-a-kinase
6	271.5	8.0	2230	COM4 HUMAN	Q13439 homo sapien
7	269	7.9	1979	TRIA HUMAN	Q15643 homo sapien
8	266.5	7.8	2663	CENE HUMAN	Q02224 homo sapien
9	266	7.8	944	NUP1 YEAST	P32380 saccharomyc
10	262.5	7.7	1690	C190 DROME	Q9vj55 drosophila
11	262	7.7	1961	MYH9 RAT	Q62812 rattus norv
12	261.5	7.7	1972	MYH9 RABIT	P35748 oryctolagus
13	260.5	7.7	1411	EEAL HUMAN	Q15075 homo sapien
14	258	7.6	1972	MYH8 MOUSE	C08638 mus musculus
15	256.5	7.5	1790	USO1 YEAST	P25386 saccharomyc
16	255	7.5	1978	MYH8 CHICK	P10587 gallus gall
17	254	7.5	2245	MY5J DICDI	P54697 dictyosteli
18	253	7.4	1939	MYH6_NESAU	P13539 mesocricetu

19	252.5	7.4	1960	1	MYH9 HUMAN	P35579 homo sapien
20	251.5	7.4	2442	1	CEP2 HUMAN	Q9bv73 homo sapien
21	250.5	7.4	1939	1	MYH6 HUMAN	P13533 homo sapien
22	250	7.3	790	1	EEAL MOUSE	Q8b166 mus musculus
23	249	7.3	1938	1	MYH6 RAT	P02563 rattus norv
24	247	7.3	1938	1	MYH6 MOUSE	Q02566 mus musculus
25	247	7.3	1976	1	MYH8 HUMAN	P35580 homo sapien
26	247	7.3	2238	1	GOA4 MOUSE	Q91yw5 mus musculus
27	247	7.3	3210	1	CENF HUMAN	P49454 homo sapien
28	247	7.3	3259	1	GOBI HUMAN	Q14789 homo sapien
29	246	7.2	1976	1	MYH8 BOVIN	Q27991 bos taurus
30	244	7.2	879	1	RA50 SULTO	Q96vrs sulfolobus
31	244	7.2	1130	1	VL17 CAEEL	Q11102 caenorhabdi
32	243.5	7.2	1935	1	MYH7 HUMAN	P12983 homo sapien
33	243	7.1	1364	1	KTNI_CHICK	Q90831 gallus gall
34	243	7.1	1935	1	MYSS CYPCA	Q90339 cyprinus ca
35	241.5	7.1	1360	1	CING XENLA	Q9ptd7 xenopus lae
36	241	7.1	997	1	SCPI RAT	Q03410 rattus norv
37	240.5	7.1	1427	1	REST HUMAN	P30622 homo sapien
38	240	7.1	1020	1	CF60 HUMAN	Q8nb25 homo sapien
39	240	7.1	1972	1	MYH8 HUMAN	P35749 homo sapien
40	240	7.1	1976	1	MYH8 RAT	Q9jito rattus norv
41	239.5	7.0	993	1	SCPI MOUSE	Q62209 mus musculus
42	238.5	7.0	1163	1	SBCC CLOAB	Q97fk1 clostridium
43	238	7.0	1938	1	MYHD HUMAN	Q9ukx3 homo sapien
44	236.5	6.9	978	1	RA50 AQUAE	C67124 aquifex aeo
45	236.5	6.9	1935	1	MYH7 RAT	P02564 rattus norv

ALIGNMENTS

RESULT 1

AN18 HUMAN	STANDARD;	PRT;	992 AA.
AC Q81v16; Q7Z468;			
DT 15-MAR-2004 (Rel. 43, Created)			
DT 15-MAR-2004 (Rel. 43, Last sequence update)			
DT 15-MAR-2004 (Rel. 43, Last annotation update)			
DE Ankyrin repeat domain protein 18A.			
GN ANKRD18A OR KIAA2015.			
OS Homo sapiens (Human)			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE=Brain;			
RA Nagase T., Kikuno R., Ohara O.;			
RT "The nucleotide sequence of a long cDNA clone isolated from human.";			
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.			
RN [2]			
RP SEQUENCE OF 1-349 FROM N.A.			
RC TISSUE=Prostate;			
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,			
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,			
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,			
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,			
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,			
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;			
RT "Generation and initial analysis of more than 15,000 full-length			
human and mouse cDNA sequences.";			

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC -/- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB095935; BAC23111.1; ALT INIT.
DR EMBL; BC056266; AAH56266.1; ALT_TERM.
DR Genew; HGNC:23643; ANKRD18A.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 6.
DR PRINTS; PR01415; ANKVRIN.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PSS0088; ANK_REPEAT; 4.
DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat.
FT REPEAT 67 96 ANK 1.
FT REPEAT 100 129 ANK 2.
FT REPEAT 133 162 ANK 3.
FT REPEAT 166 195 ANK 4.
FT REPEAT 199 228 ANK 5.
FT CONFLICT 130 130 K -> E (IN REF. 2).
FT CONFLICT 288 288 MISSING (IN REF. 2).
SQ SEQUENCE 992 AA; 115664 MW; B9B28F087340D9A CRC64;

Alignment Scores:
Pred. No.: 1,37e-16 Length: 992
Score: 443.50 Matches: 131
Percent Similarity: 50.84% Conservative: 80
Best Local Similarity: 31.57% Mismatches: 164
Query Match: 13.03% Indels: 40
DB: 1 Gaps: 6

US-09-602-362E-15 (1-2030) x AN18_HUMAN (1-992)
QY 815 ACAATGAAAGAAAGCTTTGGTACTGTAAGAAAGAACTGTGAGA----- 859
Db ThrAlaGluAspTyrAlaLeuCySerAspLeuArgSerIleArgGlnGlnIleLeuGlu 254
QY 860 -----AGCAAGAAATAAATACAGTTAGACACCAAAAGTTAAATGGACACAGA 913
Db HisLysAsnLysMet-LeuLysAsnHisLeuArgAsnAsp----- 267
QY 914 GCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAAATGCCGATAT 973
Db -----AsnGlnGluThrAlaAlaMetLysProAla----- 277
QY 974 ATTAATGAAAGAAATAGGAGAGAAATAGGAGAGATCGAGAGACGATAGGAAAGATT 1033
Db -----AsnLeuLysLysArgLysGluArgAlaLysAlaGluHisAsnLeuLysValaLase 296
QY 1034 AGAAGTGAACCAACACTTGAACAGCTCTCAGAATACAGAATATAGAAATGAAGAGTGT 1093
Db rGluGluLysGlnGlu-----ArgLeuGlnArgSe 306
QY 1094 AGAAATTAATGGAATCAGTTTCTCAGCTCATGAAATGAATGAATATCTCTTACATGA 1153
Db rGluAsnLysGlnProGlnAspSerGlnSerTyrGlyLysLysLysAspAlaMetTyrG1 326
QY 1154 AAATTCATGTTGAAAAGCAATTCCTCATGTAATAAAGTGAAGTGAAGTGAAGTGAAGT 1213
Db yasnPheMetLeuLysLysAspIleAlaMetLeuLysGluGluLeuTyrAlaIleLysAs 346
QY 1214 CCAATACCAAGGAAAGAAATTAATTAATCTTTGAGGACATTAAGATTAAAGAAAGAA 1273
Db nAspSerLeuArgLysGluLysLysTyrIleGlnGluIleLysSerIleThrGluIleAs 366
QY 1274 TGCTGAATTCAGATGACCTTAAACCTGAAGAGGAATCATTAACATAAAGGGCATCTCA 1333

Db nAlaAsnPheGluLysSerValArgLeuAsnGluLysMetIleThrLysThrValAlaAr 386
QY 1334 ATATAGTGGCAGCTTAAAGTTCTGATAGCTAGACACACATCTCCTACTCTCTAAATG-- 1391
Db gTyrSerGlnGlnLeuAsnAspLeuLysAlaGluAsnAlaArgLeuAsnSerGluLeuG1 406
QY 1392 -AAGGAAAAACAAGACAAAGAAATATCTAGAGGAGAGAAATGAATCAACCATCTAGACT 1450
Db uLysGluLysHisAsnLysGluArgLeuGluAlaGluValGluSerLeuHisSerSerLe 426
QY 1451 GCCTTCCTGCTGTACAGACCATCATCAATTGTGACATCAAGAAAAAGTCAAGAACTGC 1510
Db uAlaThrAlaIleAsnGluTyrAsnGluIleVal---GluArgLysAspLeuGluLeuVa 445
QY 1511 TTTCCACATTCGAGGAGATGCTTTGTCACAAAGAAATGAATGTTGATGTGATAGTAGTAC 1570
Db lLeuTyrArgAlaAspAspValSerArgHisGluLysMetGlySerAsnIleSerGlnLe 465
QY 1571 CGATATATAAATGAGGTCTCTCAATTCAGAGGAGATGCTCTAAGAGAAATATACATTGTTTCAG 1630
Db u-ThrAspLysAsnGluLeuThrGluGlnValHisLysAlaArgValLysPheAsnT 485
QY 1631 GCCTAAAAATTAATCTCAATTCAGAGGAGATGCTCTAAGAGAAATATACATTGTTTCAG 1690
Db hrLeuLysGlyLysLeuArgGluThrArgAspAlaLeuArgGluLysThrLeuAlaLeuG 505
QY 1691 GAACATGCACAAAGAGACCAACCTGAAACACAGTGTCAATGAAGAAAGTCAACACATG 1750
Db lSer-ValGlnLeuAspLeuArgGlnAlaGlnHisArgIleLysGluMetLysGlnMet 524
QY 1751 TTTCAANCGAACAGATNATGTGAACAAACACACTGANCAGCAGGAGTCTCTAGATCAG 1810
Db HisProAsnGlyGluAlaLysGluSerGlnSerIleGlyLysGlnAsnSerLeuGluGlu 544
QY 1811 AAATTTATTTCAACTCAACAGCAAAATATGTGCTTCAACAGCAATTAGTTCATGCACAT 1870
Db ArgIleArgGlnGlnGluLeuGluAsnLeuLeuGluArgGlnLeuGluAspAlaArg 564
QY 1871 AANGAAGCTGCACACAAAGCAAGATACAAATGATNTTCATNTCTTGAGAGGAAAT 1930
Db LysGluGlyAspAsnLysGluIleValIleAsnIleHisArgAspCysLeuGluAsnGly 584
QY 1931 GCNCATCATCTTCTAAAGAGAGAAATAGAGAGATATTTNATACNATAACCATTTAAAA 1990
Db LysGluAspLeuLeuGluGluArgAsnLysGluLeuMetLysGluTyrAsnTyrLeuLys 604
QY 1991 AACCGGTATTTCAATATGGAAGAAAAAANAAAAA 2029
Db GlulysLeuLeuGlnCysGluLysGluLysAlaGluArg 617

RESULT 2
ID MLPI_YEAST STANDARD; PRT; 1875 AA.
AC Q02455;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin-like protein MLP1.
GN MLP1 OR YKR095W OR YKR415.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID:4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=33247549; PubMed=8483450;
RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
RT "A new yeast gene with a myosin-like heptad repeat structure.";
RL Mol. Gen. Genet. 237:359-369 (1993).
RN [2]
RP SEQUENCE FROM N.A.


```

CC meromyosin (IMM) and 1 heavy meromyosin (HMM). It can be further
CC split into 2 globular subfragments (S1) and 1 rod-shaped
CC subfragment (S2).
CC
CC -! DOMAIN: The rodlike tail sequence is highly repetitive, showing
CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
CC characteristic for alpha-helical coiled coils.
CC
CC -! PTM: Phosphorylation inhibits thick filament formation and reduces
CC the actin-activated ATPase activity.
CC
CC -! MISCELLANEOUS: Dictyostelium myosin II has no K(2)EDTA ATPase
CC activity, perhaps correlated with the absence of a Cys at the SH-1
CC position (688).
CC
CC -! SIMILARITY: Contains 1 myosin-like globular head domain.
CC
CC -! SIMILARITY: Contains 1 IQ domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; M14628; AAA33227.1; -
CC
CC PIR; A26655; A26655.
CC
CC PDB; 1MMA; 03-DEC-97.
CC
CC PDB; 1MMD; 17-AUG-96.
CC
CC PDB; 1MMG; 03-DEC-97.
CC
CC PDB; 1MMN; 03-DEC-97.
CC
CC PDB; 1MND; 17-AUG-96.
CC
CC PDB; 1MNE; 17-AUG-96.
CC
CC PDB; 1VOM; 23-DEC-96.
CC
CC PDB; 1LVK; 28-JAN-98.
CC
CC PDB; 1DXX; 20-DEC-00.
CC
CC PDB; 1DOY; 20-DEC-00.
CC
CC PDB; 1DOZ; 20-DEC-00.
CC
CC PDB; 1DIA; 20-DEC-00.
CC
CC PDB; 1DIB; 20-DEC-00.
CC
CC PDB; 1DLC; 20-DEC-00.
CC
CC PDB; 1FMV; 20-DEC-00.
CC
CC PDB; 1FMW; 20-DEC-00.
CC
CC PDB; 1G8X; 17-JAN-01.
CC
CC PDB; 1JWY; 07-NOV-01.
CC
CC PDB; 1JX2; 07-NOV-01.
CC
CC DictyBase; DDB0002015; mhca.
CC
CC InterPro; IPR000048; IQ_region.
CC
CC InterPro; IPR001609; myosin_head.
CC
CC InterPro; IPR004009; Myosin_N.
CC
CC InterPro; IPR008989; Myosin_S1_N.
CC
CC Pfam; PF00612; IQ; 2.
CC
CC Pfam; PF00063; myosin_head; 1.
CC
CC Pfam; PF02736; Myosin_N; 1.
CC
CC PRINTS; PRO1193; MYOSINHEAVY.
CC
CC ProDom; PD000355; myosin_head; 1.
CC
CC SMART; SM00015; IQ; 1.
CC
CC SMART; SM00242; MYSC; 1.
CC
CC PROSITE; PS00096; IQ; 1.
CC
CC Myosin; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
CC Calmodulin-binding; Methylation; Phosphorylation.
CC
CC DOMAIN 1 761 MYOSIN HEAD-LIKE.
CC
CC DOMAIN 762 791 IQ.
CC
CC DOMAIN 817 2116 COILED COIL (POTENTIAL).
CC
CC NP_BIND 179 186 ATP.
CC
CC DOMAIN 638 660 ACTIN-BINDING.
CC
CC DOMAIN 738 752 ACTIN-BINDING.
CC
CC MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
CC
CC MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
CC
CC MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
CC
CC MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
CC
CC TURN 3 5
CC
CC TURN 7 8
CC
CC TURN 10 15
CC
CC HELIX 25 28
CC
CC STRAND 34 37

```

```

FT TURN 41 42
FT STRAND 48 55
FT STRAND 59 63
FT STRAND 69 73
FT HELIX 74 76
FT STRAND 78 79
FT HELIX 83 85
FT TURN 86 87
FT STRAND 90 90
FT HELIX 91 93
FT HELIX 99 110
FT TURN 111 113
FT STRAND 116 119
FT TURN 120 121
FT STRAND 122 126
FT HELIX 137 142
FT TURN 143 145
FT HELIX 148 150
FT HELIX 155 169
FT TURN 170 170
FT STRAND 173 179
FT TURN 181 182
FT HELIX 185 200
FT HELIX 210 226
FT STRAND 227 228
FT TURN 234 235
FT STRAND 236 237
FT STRAND 240 247
FT TURN 249 250
FT STRAND 253 261
FT HELIX 265 268
FT TURN 269 269
FT TURN 273 274
FT STRAND 278 278
FT HELIX 279 287
FT TURN 290 296
FT TURN 297 297
FT HELIX 301 303
FT TURN 305 307
FT TURN 316 317
FT HELIX 320 334
FT TURN 335 335
FT HELIX 338 355
FT TURN 356 356
FT STRAND 360 360
FT STRAND 368 368
FT HELIX 373 382

```

Alignment Scores:

```

Pred. No.: 4,56e-08 Length: 2116
Score: 280.00 Matches: 158
Percent Similarity: 36.61% Conservative: 110
Best Local Similarity: 21.58% Mismatches: 256
Query Match: 8.23% Indels: 208
DB: 1 Gaps: 27

```

US-09-602-362E-15 (1-2030) x MYO2_DICD1 (1-2116)

```

Qy 24 CTGAGCGCTAACTGCGGAATGAAGATTCTATTCCAACTAAAGCCTTAGAATTG---ATG 80
    |||||
Db 879 LeuLysAlaMetTyrAspSerLysAspAlaLeuGluAlaGlnLysArgGluLeuGluLeu 898
    |||||
Qy 81 GACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGCCATCTCGCTTCGAGCCTGCCATT 140
    |||||
Db 899 ArgValGluAspMetGluSerGluLeuAspGluLysLeuAlaLeuGluAsnLeuGln 918
    |||||
Qy 141 GAAATGCAAAAGTCTGTTCAAATAAAGCCCTTG-----GAATTGAAGATGAA 188
    |||||
Db 919 AsnGlnLysArgSerValGluGluLysValArgAspLeuGluGluLeuGlnGluGlu 938
    |||||
Qy 189 CAAACATTGAGACGACGATGAGATACCTCCCATCAGAAATCCAAACAAAGGACTATGAAGAA 248
    |||||
Db 939 GlnLysLeuArgAsn-----ThrLeuGluLysLysLysLysTyrGluGlu 954

```

QY 249 AGTTCTTGGGAT-----TCTGAGAGTCTCTGTGAGACTGTCTTTCACAGAAAGGAT 296
 Db 955 GluLeuGluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGlu 974
 QY 297 GTGTGTTTACCCAAAGCTCACATCAAAAGAAATAGATAAAATAAATGGAATAA--- 353
 Db 975 LysIle-----LysAspGlnLeuGlnLysGluValGluLeuThrGlnSerPheSer 992
 QY 354 GAAGAGTCTCTGATTAATGATGTTTCTGAGAGCTCCCTCGCAGAAATGAAAGTTCTATT 413
 Db 993 GluGluSerLysAspLysGlyValLeuGluLysThrArgValArgLeuGlnSerGluLeu 1012
 QY 414 CCAACTAAAGCCTTAGAATGTGGAGCATGCAAACTTTCAAAGCAGAG----- 461
 Db 1013 AspAspLeuThrValArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGln 1032
 QY 462 -----CCTCCGAGAGCCATCTGCCCTCGAGCTGCC 494
 Db 1033 LysLysLysLeuGluGluLeuLysGlnValGlnGlnAlaLeuAlaGluThrAla 1052
 QY 495 ATTGAATGCCAAAGCTCTGTTCCAAATAAAGCCTTG----- 530
 Db 1053 AlaLysLeuAlaGlnGluAlaAlaAsnLysLysLeuGlnGlyGlyThrThrGluLeuAsn 1072
 QY 530 ----- 530
 Db 1073 GluLysPheAsnSerGluValThrAlaArgSerAsnValGluLysSerLysLysThrLeu 1092
 QY 531 -----CAATTGAAGNATCAACAACA 551
 Db 1093 GluSerGlnLeuValAlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAla 1112
 QY 552 TTGAGA-----CCAGATCAGATGTC 572
 Db 1113 LeuGluLysLysLysLysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGlu 1132
 QY 573 CCTTCAGATCAAAACAAAG-----AAGTTGAAGAAATCTTGGGAT 617
 Db 1133 SerThrGlyGluLysLysSerLeuTyArgLeuLysValLysGlnGluSer---Asp 1151
 QY 618 TCTGAGAGTCTCGTGAGAGTCTTTCACAGAGGATGTGTGTGCC----- 665
 Db 1152 MetGluAlaLeuArgGlnIleSerGluLeuGlnSerThrIleAlaLysLeuGluLys 1171
 QY 666 ---AAGGCTACACATCAAAAGAAATGGATAAAATAGTGGAAATTAAGAGATTCAACT 722
 Db 1172 IleLysSerThrLeuGluGlyGluValAlaArgLeuGlnGlyGluLeuGlu----- 1188
 QY 723 AGCCTATCAAAATCTTGATACAGTTCATTCTGTGAAGAGCAAGGAACTTCAAAA 782
 Db 1189 -----AlaGlnLeuAlaLys 1194
 QY 783 GATCACTGTGACACAGCTACAGAAATCGAA---CAAATGAAAGAAAGTTTTGTGTA 839
 Db 1195 SerAsnValGluLysGlnLysLysLysValGluLeuAspLeuGluAspLys----- 1211
 QY 840 CTGAAAAGAACTGTGACAGCAAGCAAGAAATAAATACAGTTAGACNACCAAAAGTT 899
 Db 1212 ---SerAlaGlnLeuAlaGluGluThrAlaAlaLysGlnAlaLeuAspLysLeuLysLys 1230
 QY 900 AAATGGGAACAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGA 959
 Db 1231 LysLeuGluGlnGluLeuSerGluValGlnThrGlnLeuSerGluAlaAsnAsnLysAsn 1250
 QY 960 AGAAATGCCGATATATAATGAAATAATAGGAAAGAAATAGGAAAGATTCGAGAGCAG 1019
 Db 1251 ValAsnSerAspSerThrAsnLysHisLeuGluThrSerPheAsnAsnLeu----- 1267
 QY 1020 CATAGGAAGAGTGTAGAGTGAACCAACCTTGAACAGGCTCTCAGATATCAAGATATA 1079
 Db 1268 ---LysLeuGluLeuGluAlaGluGlnLysAlaLysGlnAlaLeuGluLysLysArgLeu 1286

QY 1080 GAATTGAAGAGTGTAGAAAGTAATTTGAATCAAGTTTCTCACACTCATGAAATGAAAT 1139
 Db 1287 GlyLeu-----GluSerGluLeuLysHisValAsnGluGlnLeuGluGluLys 1303
 QY 1140 TATCTCTTACATGAAATTCATGTTGAAAGAAATGCTCCATGCTTAAACTGGAATA 1199
 Db 1304 LysGlnLysGluSerAsn-----GluLysArgLysValAspLeuGluLysGluVal 1320
 QY 1200 GCCACACTGAAACACCAATACACAGGA----- 1226
 Db 1321 SerGluLeuLysAspGlnIleGluGluGluValAlaSerLysLysAlaValThrGluAla 1340
 QY 1227 -----AAGGAAAT-----AAATACTTTGAGGACATTAAGATT 1259
 Db 1341 LysAsnLysLysGluSerGluLeuAspGluIleLysArgGlnTyrAlaAspValValSer 1360
 QY 1260 TTAAGAAAGAAATGCTGAACTTCAGATGACCTTAAACTGAAAGAGAAATCATTAAT 1319
 Db 1361 SerArgAspLysSerValGluGlnLeuLysThrLeuGlnAlaLysAsnGluGluLeuArg 1380
 QY 1320 AAAGAGGCACTCTCAATATAGTGGCAGCTTAAAGTTCTCATAGCTGAGAACACATGCTC 1379
 Db 1381 AsnThrAlaGluGluAlaGluGlyGlnLeu----- 1390
 QY 1380 ACTTCTTAATTTGAAGAAACAAAGCAAGAAATACTAGAGGCAGCAAAATGATCACAC 1439
 Db 1391 -----AspArgAlaGluArgSerLysLysLysAlaGluPheAsp----- 1403
 QY 1440 CATCTAGACTGCTTCTGCTGTACAAAGCAACCATCAAAATTTGACATCAAGAAAAAGT 1499
 Db 1404 -----LeuGluGluAlaValLysAsnLeuGluGluThrAlaLysLysValLys 1420
 QY 1500 CAAGAACTCTCTTCCACATTCGAGGAGATGTTGTTGCAAGAAAAATGAATGTTGAT 1559
 Db 1421 AlaGluLysAla-----MetLysLysAlaGluThrAsp 1431
 QY 1560 GTGAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAA 1619
 Db 1432 TyrArgSerThr-----LysSer 1437
 QY 1620 GAAATCCAAAGCCTTAAATTAATCTCAATTATGAGGAGATGCTTAAAGAAAAATAC 1679
 Db 1438 GluLeuAspAspAlaLysAsn-----ValSerSerGluGlnTyr 1450
 QY 1680 ATTGTTTTCAGGAACATGCAAAAGAGACCAACGCTGAAACACAGTCTCAATGAAGGA 1739
 Db 1451 ValGlnIleLysArgLeuAsnGluGluLeuSerGluLeuArgSerValLeuGluAla 1470
 QY 1740 CTGAACACATGTTTCAAAACGAAAGATNATGTGAACAAACACACTGANCACAGGAGT 1799
 Db 1471 AspGluArgCysAsnSerAlaIleLysAlaLysLysThrAlaGluSerAlaLeuGluSer 1490
 QY 1800 CTCTAGATCAGAAATTTATTTCAACTCAAGCAAGCAAAATATGTGCTTCAACACCAATTAG 1859
 Db 1491 Leu-----LysAspGluIleAspAlaAlaAsnAsn----- 1500
 QY 1860 TTCATGCACATAANGAAAGCTGACAAACAAAGCAAG 1895
 Db 1501 -----AlaLysAlaLysAlaGluArgLysSerLys 1510
 RESULT 4
 MYSN DROME
 ID MYSN DROME STANDARD; PRT; 2017 AA.
 AC Q99323;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 294, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, non-muscle (zipper protein) (Myosin II).
 GN ZIP.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

Db 1460 AspArgLeuAspLysSerLysLysLysIleGlnSerGluLeuGluAspAlaThrIleGlu 1479
Qy 903 TGGGAACAGAGCTCTGCAGCTGTGACATTCATTAAACCAAGACAGAGAGAGA 962
Db 1480 LeuGluAlaGln-----ArgThrLysValLeuGluLeuGluLysGlnLys 1495
Qy 963 AATGCCGAT-----ATATTAAATGAATAATAGGGAAGATTA 1001
Db 1496 AsnPheAspLysIleLeuAlaGluGluLysAlaIleSerGluGlnIleAlaGlnArg 1515
Qy 1002 GGAAGATCGAA-----GAGCAGCATAGGAAGAGTTAGAGTGAACCAACA 1049
Db 1516 AspThrAlaGluArgGluAlaArgGluLysGluThrLysValLeuSerValSerArgGlu 1535
Qy 1050 CTTGAACAGGCTCTC---AGAAATACAGATATAGAA-----TTGAAGAGTGTAGAAAGT 1100
Db 1536 LeuAspGluAlaPheAspLysIleGluAspLeuGluAsnLysArgLysThrLeuGlnAsn 1555
Qy 1101 AATTGAATCAGTTCTCAGCTCATGAAATGAATAATATCTCTCATCATGAAATGTC 1160
Db 1556 GluLeuAspLeuAlaAsnThrGlnGlyThrAlaAspLysAsnValHisGlu----- 1573
Qy 1161 ATGTTGAAAAGAAATTCGCTGCTAAACTGGAATAGCCACACTGAAACCAACATAC 1220
Db 1574 ---LeuGluLysAlaLysArgAlaLeuGluSerGlnLeuAlaGluLeuLysAlaGlnAsn 1592
Qy 1221 CAGGAAAAGAAAT-----AAATCTTTGAGGACATTAAGATTTTAAAGAAAAGAT 1274
Db 1593 GluGluLeuGluAspLeuGlnLeuThrGluAspAlaLysLeuArgLeuGluValAsn 1612
Qy 1275 -----CCTGAATCTCAGATGACCCCTAARACTGAAGAGAGAAATTAATCT 1319
Db 1613 MetGlnAlaLeuArgSerGlnPheGluArgAspLeuLeuAlaLysGluGluGlyAlaGlu 1632
Qy 1320 AAAAGGCGATCAATATATAGTGGCAGCTTAAAGTTCTGATAGCTGAG----- 1367
Db 1633 GluLysArgArgLysLeuValLysGlnLeuAspLeuGluThrGluLeuAspGluGlu 1652
Qy 1368 -----AACCAATGCTCACTTCTTAATGAAGGAAACACAGACAAAGATA 1415
Db 1653 ArgLysGlnArgThrAlaAlaValAlaSerLysLysLeuGluGluGlyAspLeuLysGlu 1672
Qy 1416 CTAGAGCGAGAATTTCAATCACACACCTAGCTGGCTTCTGCTGTACAGACCATCAT 1475
Db 1673 IleGluThrThrMetGluMethIleAsnLysValLysGluAspAlaLeu-LysHisAlaLys 1692
Qy 1476 CAAATGTGACATCAAGAAAGTCAAGAACCTGCTTTCCACATTCGAGGAGATCCTGT 1535
Db 1692 sLysLeu-----GlnAlaGlnValLysAspAlaLeuArgAspAlaGluGluAlaLysAl 1710
Qy 1536 TTGCAAGAAAATGAATGTTGATG-----TCAGTAGTACCGATATATAA 1580
Db 1710 alaLysGluGluLeuGlnAlaLeuSerLysGluAlaAspGlyLysValLysAlaLeuGlu 1730
Qy 1581 CAATGAGTGTCTCAT-----CAACCATTTCTGAAGCTCAAGGAAATCCANAGCCT 1634
Db 1730 uAlaGluValLeuGlnLeuThrGluAspLeuAlaSerSerGluArgAlaArgAlaAl 1750
Qy 1635 AAAAATTAATCTCAATTATGCAGGAGATGCTCTAAGAGAAATACATTTGTTTCAGGAAC 1694
Db 1750 aGluThrGluArgSpGluAlaGluGluIleAlaIleAlaIleAlaIleAlaIleAlaIle 1770
Qy 1695 ATGCACAAAGAGACCAACGCTGAACACAGCTGCAATGAAGGAAAGCTGAACACATGTTTC 1754
Db 1770 uMetIleAsp-GluLysArgArgLeuGluAlaIleAlaIleAlaIleAlaIleAlaIle 1790
Qy 1755 AAANCAACAAGAT-----NATGTGAACAACACACTGANCAGCAGGAGTCTC 1802
Db 1790 luGluGluGlnSerAsnSerGluValLeuLeuAspArgAlaAlaAlaAlaAlaAlaGlnIle 1810
Qy 1803 TAGATCAGAAATATTATTCAACTCAAGCAAGAAATATGTGGCTTCAACAGCAATAGTTC 1862
Db 1810 leGluGlnLeuThrThrGluLeu----- 1817

Qy 1863 ATGCACATANGAAAGCTGACACAAAGCAAGATTAATGATTTTCATTTCCTTGTAG 1922
Db 1818 --AlaAsnGluLysSerAsnSerGlnLys-----A 1827
Qy 1923 AGGAAATGNCATCATCTTCTTAAAGAGAGAAATAGAGAGATA 1966
Db 1827 snGluAsnGlyArgAlaLeuLeuGluArgGlnAsnLysGluLeu 1841
RESULT 5
AKA9_HUMAN
ID AKA9_HUMAN STANDARD; PRT; 3911 AA.
AC Q99996; Q14869; O43355; O94895; Q9UQH3; Q9UQQ4; Q9Y6B8; Q9Y6Y2;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
PKN-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98151389; PubMed=9482789;
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RT "Yotiao, a novel protein of neuromuscular junction and brain that
interacts with specific splice variants of NMDA receptor subunit
NRI.";
RL J. Neurosci. 18:2017-2027 (1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,
RA Janssen T., Oerstavik S.;
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
protein located in the centrosome, AKAP450.";
RL EMO J. 18:1858-1868 (1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
anchors multiple signaling enzymes to centrosome and the Golgi
apparatus.";
RL J. Biol. Chem. 274:17267-17274 (1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX Kemmer W.A., Deiss S., Schwarz U.;
RT "Cloning of Hyperion.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915845;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RA Trotter K.W., Milgram S.L., Goldenring J.R.;
RT "AKAP350, a multiply spliced protein kinase A-anchoring protein
associated with centrosomes.";
RL J. Biol. Chem. 274:3055-3066 (1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RX Hinds K., Sutterer C., Becker M., Hawkins M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).

RC TISSUE=Lung;
RA Milgram S.L., Goldemring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Isikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Korani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT for large proteins in vitro.";
RT The complete sequences of 100 new cDNA clones from brain which code
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds to type II regulatory subunits of protein kinase
CC A. Scaffolding protein that assembles several protein kinases and
CC phosphatases on centrosome and Golgi apparatus where physiological
CC events can be regulated by phosphorylation state of protein
CC substrates. Isoform 4/Yotiao is associated with the N-methyl-D-
CC aspartate receptor and is specifically found in the neuromuscular
CC junction (NMJ) as well as in neuronal synapses explaining that its
CC role may be to organize postsynaptic specializations.
CC -!- SUBUNIT: Interacts with the regulatory region of protein kinase N
CC (PKN), protein phosphatase 2A (PP2A), protein phosphatase 1 (PP1)
CC and the immature non-phosphorylated form of PKC epsilon.
CC -!- SUBCELLULAR LOCATION: Centrosomal in many cell types and
CC cytoplasmic in parietal cells.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=1;
CC IsoId=Q99996-1; Sequences=Displayed;
CC Name=2;
CC IsoId=Q99996-2; Sequences=VSP_004102, VSP_004107;
CC Name=3;
CC IsoId=Q99996-3; Sequences=VSP_004102, VSP_004105, VSP_004107;
CC Name=4;
CC IsoId=Q99996-4; Sequences=VSP_004103, VSP_004104;
CC Name=5;
CC IsoId=Q99996-5; Sequences=VSP_004108;
CC Name=6;
CC IsoId=Q99996-6; Sequences=VSP_004106, VSP_004107, VSP_004109;
CC TISSUE SPECIFICITY: Widely expressed. Isoform 4/Yotiao is highly
CC expressed in skeletal muscle and in pancreas.
CC -!- DOMAIN: RII-binding site, predicted to form an amphipathic helix,
CC could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer.
CC -!- CAUTION: Ref.6 sequence differs from that shown due to two
CC frameshifts in positions 3782 and 3811.
CC -!- CAUTION: Ref.9 sequence differs from that shown due to four
CC frameshifts in positions 29, 1653, 1699 and 1735.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ131693; CAB40713.1; -;
CC EMBL; AB019691; BAA78718.1; -;
CC EMBL; AJ010770; CAA09361.1; -;
CC EMBL; AF026245; AAB86384.1; -;
CC EMBL; AF083037; AAD22767.1; -;
CC EMBL; AC004013; AAB96867.1; ALT_FRAME.
CC EMBL; AF091711; AAD39719.1; -;
CC EMBL; AB018346; BAA34523.1; -;
CC EMBL; AC000066; AAC60380.1; ALT_FRAME.

DR Genew; HGNC:379; AKAP9.
DR MIM; 604001; -;
DR GO; GO:0005813; C:cytoskeleton; TAS.
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr...; TAS.
DR GO; GO:000515; F:protein binding; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
DR GO; GO:0006810; P:transport; TAS.
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 914 COILED COIL (POTENTIAL).
FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT DOMAIN 3587 3689 COILED COIL (POTENTIAL).
FT DOMAIN 3726 3730 POLY-LEU.
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 GLU-RICH.
FT VARSPLIC 17 28 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 1637 1642 /FTid=VSP_004102.
FT VARSPLIC 1643 3911 QLQEEI -> LATRD (in isoform 4).
FT VARSPLIC 2175 2182 /FTid=VSP_004103.
FT VARSPLIC 2175 2182 Missing (in isoform 4).
FT VARSPLIC 2175 2182 /FTid=VSP_004104.
FT VARSPLIC 2175 2182 Missing (in isoform 3).
FT VARSPLIC 2175 2183 /FTid=VSP_004105.
FT VARSPLIC 2895 2907 SADTFQKVE -> Q (in isoform 6).
FT VARSPLIC 2895 2948 /FTid=VSP_004106.
FT VARSPLIC 2895 2948 VFGFVNMCFSTLC -> GSSIPBLAHSDAYQTRICSS
FT VARSPLIC 3901 3911 (in isoform 2, isoform 3 and isoform 6).
FT VARSPLIC 3901 3911 /FTid=VSP_004107.
FT VARSPLIC 3901 3911 Missing (in isoform 5).
FT VARSPLIC 3901 3911 /FTid=VSP_004108.
FT VARSPLIC 3901 3911 STTQFAGMER -> ALSLTTSWQHSARPTAPLFFELSH
FT VARSPLIC 3901 3911 SLG (in isoform 6).
FT VARSPLIC 3901 3911 /FTid=VSP_004109.
FT VARSPLIC 3901 3911 K -> Q.
FT VARSPLIC 3901 3911 /FTid=VAR_010926.
FT VARSPLIC 3901 3911 E -> Q (IN REF. 3).
FT VARSPLIC 3901 3911 M -> I (IN REF. 3).
FT VARSPLIC 3901 3911 E -> G (IN REF. 3).
FT VARSPLIC 3901 3911 R -> S (IN REF. 3).
FT VARSPLIC 3901 3911 N -> S (IN REF. 3).
FT VARSPLIC 3901 3911 H -> N (IN REF. 3).
FT VARSPLIC 3901 3911 K -> N (IN REF. 3).
FT VARSPLIC 3901 3911 QKH -> PKP (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 Q -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 N -> D (IN REF. 3).
FT VARSPLIC 3901 3911 V -> E (IN REF. 3).
FT VARSPLIC 3901 3911 R -> P (IN REF. 1 AND 2).
FT VARSPLIC 3901 3911 V -> T (IN REF. 3).
FT VARSPLIC 3901 3911 V -> G (IN REF. 3).
FT VARSPLIC 3901 3911 MISSING (IN REF. 5).
FT VARSPLIC 3901 3911 A -> P (IN REF. 3).
FT VARSPLIC 3901 3911

Alignment Scores:

Pred. No.: 5,05e-08
Score: 278.50
Percent Similarity: 39.52%
Best Local Similarity: 18.92%
Query Match: 8.18%
DB: 1

Length: 3911
Matches: 158
Conservative: 172
Mismatch: 250
Indels: 255
Gaps: 30

US-09-602-362E-15 (1-2030) x AKA9_HUMAN (1-3911)

135 GCCATTGAATCAAAAGCTGTTCCAAATAAAGCCTTGAATTAAGAAATGAA----- 188
112 SerValGluLeuGluSerGluIleSerThrThrAlaAspAspCysSerSerGluValAsn 131
189 -----CAAACATTGAGACGACATGAGATA 212
132 GlyCysSerPheValMetArgThrGlyLysProThrAsnLeuLeuArgGluGluPhe 151
213 CTCCCATCAGATCCAAACAAAGGACTATGAGAAAGTCTTGGAGTTCAGAGTCTC 272
152 GlyValAspSerTyrSer-----GluGlnGlyAlaGlnAspSerProThrHis 168
273 TGTGAGACTGTTTACAGAGAAGGATGTGTGTTTACCAAGCTACACATCAAAAGAAATA 332
169 LeuGluMetMetGluSerGlu-----LeuAlaGlyLysGlnHisGluIle 183
333 GATAAAATAATGAAATTAGAGAGTCTCCTGATATGATGGTTCGAAAGCTCCC 392
184 GluGluLeuAsnArgGluLeuGlu----- 192
393 TGCAGAAATGAAAGTCTTCTCACTCACTAAAGCCTTAGAATTGATGACATGCAAACTTTC 452
193 -----MetArgValThrTyrGlyThrGluGlyLeuGln-----GlnLeuGlnGluPhe 208
453 AAGCGAGAGCTCCCGAGAGGCCATCTGCCTCGAGCCTGCCATTAAGATGCAAAAGTCT 512
209 GluAlaAlaIleLysGlnArg-----AspGlyIleIleThrGlnLeuThr 223
513 GTTCCAAATAAGCCTTGGNAATTGAGAATGAACAAACATTCGACGACGATGATGTC 572
224 AlaAsnLeuGlnGlnAlaArgGluLysAspGluThrMetArgGluPheLeuGluLeu 243
573 CCTTCAGAAATCAAAACAAAGAGTTGAA-----GAAAATCTTGGGATTTCTGAGAGTCTC 629
244 ThrGluGlnSerGlnLysLeuGlnIleGlnPheGlnGlnLeuGlnAlaSerGluThrLeu 263
630 CTGAGACTGTT-----TCACAGAAGGATGTGTGTACCCCAAG-----GCT 671
264 ArgAsnSerThrHisSerThrAlaAlaAspLeuLeuGlnAlaLysGlnGlnIleLeu 283
672 ACACATCAAAAGAAATGATAAAATAAGTGAATAATTAGAAGAT----- 716
284 ThrHisGlnGlnGlnLeuGluGlnAspHisLeuLeuGluAspTyrGlnLysLysLys 303
717 -----TCAACTAGCCTATCAAAATCTTGGAT-----ACAGTTCACTTGTGAA 761
304 GluAspPheThrMetGlnIleSerPheLeuGlnGlnGluLysIleLysValTyrGluMetGlu 323
762 AGAGAAGGGAACCTTCAAAAGATCAT-----TGT 791
324 GlnAspLysLysValGluAsnSerAsnLysGluGluIleGlnGluLysGluThrIleIle 343
792 GAACAACGCTACAGGAAATGGAACAAATGAAAGAGCTTTTGTACTGGAAGAA 851
344 GluGluLeuAsnThrLysIleIleGluGluGluLysLysThrLeuGluLeuLysAspLys 363
852 CTGTCAAGAGCAAAAGAAATAAATACACAGTTAGAGAACCAAAAGTTAAATGGGAACA 911
364 LeuThrThrAlaAspLysLeuGlyGluLeuGlnGlnGlnIleValGlnLysAsnGln 383
912 GAGCTCTGAGTGTGAGATTGATTTTAAACCAAGAGAGAGAGAGAAATGCCGAT 971
384 GluIleLysAsnMetLysLeuGluLeuThrAsnSerLysGlnLysGluArgGlnSer--- 402
972 ATATTAAATCAAAATATTAGGAGAGATAGGAGNATCGAGAG----- 1016
403 -----SerGluGluIleLysGlnLeuMetGlyThrValGluGluLeuGlnLysArgAsn 420
1016 ----- 1016

421 HisLysAspSerGlnPheGluThrAspIleValGlnArgMetGluGlnGluThrGlnArg 440
1016 ----- 1016
441 LysLeuGluGlnLeuArgAlaGluLeuAspGluMetTyrGlyGlnGlnIleValGlnMet 460
1017 -----CAGCATAGGAAGAAGATTAGAGTG-----AAA 1043
461 LysGlnGluLeuIleArgGlnHisMetAlaGlnMetGluGluMetLysThrArgHisLys 480
1044 CAACAACCTTGAACACAGGCTCTCAGA----- 1067
481 GlyGluMetGluAsnAlaLeuArgSerTyrSerAsnIleThrValAsnGluAspGlnIle 500
1068 -----ATACAAGATATAGATTGAAGAGTGTAGAAATGTAATTTTGAAT 1109
501 LysLeuMetAsnValAlaIleAsnGluLeuAsnIleLysLeuGlnAspThrAsn---Ser 519
1110 CAGGTTTCTCACACTCATGAAATGAAATTTACTCTTACATGAAATTCATGTTGAAA 1169
520 GlnLysGlnLysLysGluGluLeuGlyLeuIleLeuGluGluLysCysAlaLeuGln 539
1170 AAGAAATTCCTGCTTAAACTGAAATAGCCACACTGAAACACCAATACCCAG----- 1223
540 ArgGlnLeuGluAspLeuValGluLeuSerPheSerArgGluGlnIleGlnArgAla 559
1224 -----GAAAAGGAATAATATAC----- 1241
560 ArgGlnThrIleAlaGluGlnGluSerLysLeuAsnGluAlaHisLysSerLeuSerThr 579
1242 TTTGAGGACATTAAGATTTTAAAGAAAGAAATGCTGAACTTCAGATGACCCCTAAACTG 1301
580 ValGluAspLeuLysAlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu 599
1302 AAGAGGAATCAATTAATAAGGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATA 1361
600 LysHisGluAla-----GluValThrAsnTyrLysIleLysLeuGluMetLeuGlu 616
1362 GCTGAGAACACAATGCTCACTTCTAAATTTGAAGAAACAAACAGACAAAGAAATA----- 1415
617 LysGluLysAsnAlaValLeuAspArgMetAlaGluSerGlnGluAlaGluLeuGluArg 636
1416 -----CTAGAGGAGAAAT 1430
637 LeuArgThrGlnLeuLeuPheSerHisGluGluGluLeuSerLysLysLysGluAspLeu 656
1431 GAATCACACCATCTAGACTGGCTTCTGCTGTACAAGACCATCATCAAAATTTGACATCA 1490
657 GluIleGluHisArgIleAsnIleGluLysLeuLysAspAsn-LeuGlyIleHisTyrLys 676
1491 AGAAAAGTCAAGAACCTG-----CTTTCCACATTGCG----- 1524
676 sGlnIleAspGlyLeuGlnAsnGluMetSerGlnLysIleGluThrMetGlnPheGln 696
1525 -----GAGATCTTGTTCGAAGAAATAATCAATGTTGATCTGATAGTAGTACCGATATA 1580
696 uLysAspAsnLeuIleThrLysGlnAsnGlnLeuIleLeuGluIleSerLysLeuLys 716
1581 CAATGAGGTCTCCATCAACACCTTTCTGAAGCTCAAGGAATCCANAAGCCTAAAT 1640
716 p-----LeuGlnGlnSerLeuValAsnSerLysSerGluGluMetThrLeuGlnIle 733
1641 TAAT---CTCAATATGACGAGATGCTCTAAGAGAAATAACATTTGTTTCAGGACATG 1697
733 eAsnGluLeuGlnLysGluIleGluIleLeuArgGlnGluLysGluLysGlyThr-L 753
1698 CACAAAGAGACCAACGTGAAACACAGTGT-----CAATGAAG 1736
753 euGlnGlnGluValGlnGluLeuGlnLeuLysThrGluLeuGluLysGlnMetLysG 773
1737 AGCTGAACACATGTTTCAANCGAA----- 1762
773 lylsGluAsnAspLeuGlnGluLysPheAlaGlnLeuGluAlaGluAsnSerIleLeuL 793

Db 854 hrGLuLeuAspAlaHisIysIleGlnValGlnAspLeuMetGlnGlnLeuGluIysGlnA 874

QY 1956 ATGACGAGATATTNATTACNATACCATTTTAAACCCGCTATTATTTCA 2005
|||::|
|||::|

Db 874 snSerGluMetGluGlnIysValIysSerLeuThrGlnValIysGluSer 890
|||::|
|||::|

RESULT 7

TRIA_HUMAN

ID TRIA_HUMAN STANDARD; PRT; 1979 AA.

AC Q15643; Q14689; Q15154; Q95949;

DT 01-NOV-1997 (Rel. 35, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Thyroid receptor interacting protein 11 (TRIP11) (Golgi-associated

DE microtubule-binding protein 210) (GMAP-210) (Trip230) (Clonal

DE evolution related gene on chromosome 14).

GN TRIP11 OR CEV14.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., FUNCTION, AND INTERACTION WITH RBL AND THRB.

RC TISSUE=Fibroblast;

RX MEDLINE=97404344; PubMed=9256431;

RA Chang K.-H., Chen Y., Chen T.-T., Chou W.-H., Chen P.-L., Ma Y.-T.,

RA Yang-Feng T.L., Leng L., Tsai M.-J., O'Walley B.W., Lee W.-H.;

RT "A thyroid hormone receptor coactivator negatively regulated by the

RT retinoblastoma protein.";

RL Proc. Natl. Acad. Sci. U.S.A. 94:9040-9045 (1997).

RN [2]

RP SEQUENCE FROM N.A., FUNCTION, TISSUE SPECIFICITY, SUBCELLULAR

RP LOCATION, AND INTERACTION WITH MICROTUBULES.

RC TISSUE=Cervical carcinoma;

RX MEDLINE=9207053; PubMed=10189370;

RA Infante C., Ramos-Morales F., Fedriani F., Bornens M., Rios R.M.;

RT "GMAP-210, a cis-Golgi network-associated protein, is a minus end

RT microtubule-binding protein.";

RL J. Cell Biol. 145:83-98 (1999).

RN [3]

RP SEQUENCE FROM N.A.

RX PubMed=12508121;

RA Heilig R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,

RA Cattolico L., Levy M., Barbe V., De Berardinis V., Ureta-Vidal A.,

RA Pelletier E., Vico V., Anthouard V., Rowen L., Madan A., Qin S.,

RA Sun H., Du H., Pepin K., Attiguenave F., Robert C., Cruaud C.,

RA Bruels T., Jallou O., Friedlander L., Samson G., Brottier P.,

RA Cure S., Segurens B., Aniere F., Samain S., Crespeau H., Abbasi N.,

RA Aiach N., Boscus D., Dickhoff R., Dora M., Dubois I., Friedman C.,

RA Gouyvenoux M., James R., Madan A., Mairey-Estrada B., Mangenot S.,

RA Martins N., Menard M., Oztas S., Ratcliffe A., Shaffer T., Trask B.,

RA Vacherie B., Bellemere C., Belser C., Besnard-Gonnet M.,

RA Bartol-Mavel D., Boutard M., Briez-Silla S., Combette S.,

RA Dufosse-Laurent V., Ferron C., Lechaplais C., Louesse C., Muselet D.,

RA Magdelenat G., Pateau E., Petit E., Sirvain-Trukniewicz P., Trybou A.,

RA Vega-Czarny N., Bataille E., Bluet E., Bordelais I., Dubois M.,

RA Dumont C., Guerin T., Hafray G., Hamadi R., Nwanga J., Fellouin V.,

RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,

RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., McPherson J.,

RA Matsuda F., Wilson R., Scarpelli C., Gyapay G., Winkler P., Saurin W.,

RA Quetier F., Waterston R., Hood L., Weissbach J.;

RT "The DNA sequence and analysis of human chromosome 14.";

RL Nature 421:601-607 (2003).

RN [4]

RP SEQUENCE OF 1198-1979 FROM N.A., CHROMOSOMAL TRANSLOCATION WITH

RP PDGFRB, AND TISSUE SPECIFICITY.

RC TISSUE=Leukemia;

RX MEDLINE=98043615; PubMed=9373237;

RA Abe A., Emi N., Tanimoto M., Terasaki H., Marunouchi T., Saito H.;

RT "Fusion of the platelet-derived growth factor receptor beta to a novel

RT gene CEV14 in acute myelogenous leukemia after clonal evolution.";

RL Blood 90:4271-4277 (1997).

RN [5]

RP Alignment Scores: 1.73e-07 Length: 1979

Pred. No.:

RP SEQUENCE OF 1757-1944 FROM N.A., AND INTERACTION WITH THRB.

RC TISSUE=Cervical carcinoma;

RX MEDLINE=95295737; PubMed=7776974;

RA Lee J.W., Choi H.-S., Gyuris J., Brent R., Moore D.D.;

RT "Two classes of proteins dependent on either the presence or absence

RT of thyroid hormone for interaction with the thyroid hormone

RT receptor.";

RL Mol. Endocrinol. 9:243-254 (1995).

CC -|- FUNCTION: Binds the ligand binding domain of the thyroid receptor

CC (THRB) in the presence of triiodothyronine and enhances THRB-

CC modulated transcription. Golgi auto-antigen; probably involved in

CC maintaining cis-Golgi structure.

CC -|- SUBUNIT: Binds RBL.

CC -|- SUBCELLULAR LOCATION: Peripheral membrane protein associated with

CC the cis-Golgi network. Associates with the ends of centrosome-

CC nucleated microtubules.

CC -|- TISSUE SPECIFICITY: Highly expressed in pancreas, muscle, heart,

CC testis, peripheral blood leukocytes, and in several leukemia cell

CC lines. Detected at intermediate levels in placenta and kidney, and

CC at low levels in brain and lung.

CC -|- DOMAIN: Extended rod-like protein with coiled-coil domains.

CC -|- DISEASE: Involved in a acute myelogenous leukemia through a

CC chromosomal translocation t(5;14)(q33;q32) that involves PDGFRB

CC and TRIP11. The fusion protein may be involved in clonal evolution

CC of leukemia and eosinophilia.

CC -|- SIMILARITY: Belongs to the golgin family.

CC -|- SIMILARITY: Contains 1 GRIP domain.

CC -|- CAUTION: Ref.4 sequence differs from that shown due to

CC frameshifts in position 1932 and 1955.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; AF007217; AAD09135.1; -.

DR EMBL; Y12490; CAAV3095.1; -.

DR EMBL; AL049872; -; NOT ANNOTATED CDS.

DR EMBL; AF011368; AAB84386.1; ALT FRAME.

DR EMBL; L40380; AAC41730.1; ALT_SEQ.

DR PR; T03719; T03719.

DR HSP; P04266; IIC2.

DR Genew; HGNC:12305; TRIP11.

DR MIM; 604505; -.

DR GO; GO:0005515; F:protein binding; TAS.

DR GO; GO:0003713; F:transcription co-activator activity; TAS.

DR GO; GO:0006366; P:transcription from Pol II promoter; TAS.

DR InterPro; IPR003345; W_repeat.

DR Pfam; PF02370; N; 9.

DR PROSITE; PS50913; GRIP; 1.

DR DOMAIN 52 1773

DR DOMAIN 1774 1823

DR SITE 1754 1755

FT BREAKPOINT FOR TRANSLOCATION TO FORM

FT COILED COIL (POTENTIAL).

FT GRIP.

FT TRIP11-PDGFRB

FT L -> F (IN REF. 1).

FT A -> G (IN REF. 1).

FT Q -> R (IN REF. 1).

FT D -> A (IN REF. 1).

FT KEK -> FVL (IN REF. 1).

FT H -> N (IN REF. 1 AND 4).

FT H -> R (IN REF. 4).

FT QQ -> HE (IN REF. 4).

FT QLSVSEQ -> RFLSGT (IN REF. 1).

FT A -> R (IN REF. 1).

FT S -> G (IN REF. 1).

SQ SEQUENCE 1979 AA; 227638 MW; 83B4677D34D360D CRC64;

Db 916 InLysIleileGluAspGlnAsnGlnSer-----LysMetGlnLeuLeuGlnSerL 933
Qy 1869 ATAANGAAGCTGCAACAAAGCAAGATAACAATTCATNTTCATNTCTTGAGAGGAAA 1928
Db 933 euGlnGlu-----GlnLysLysGluMetAspGluPheArgTyrGlnHisGluGlnMetA 951
Qy 1929 ATGNCATCAT-----CTTCTAAAGAGAAAATGAGGAGATA 1966
Db 951 snAlaThrHisThrGlnLeuPheLeuGluLysAspGluGluile 965
RESULT 8
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and hBUBR1.";
RL J. Cell Biol. 143:49-63(1998).
RN [4]
RP FARNESYLATION
RX MEDLINE=20459117; PubMed=10852915;
RA Ashar H.R., James L., Gray K., Carr D., Black S., Armstrong L.,
RA Bishop W.R., Kirschmeier P.;
RT "Farnesyl transferase inhibitors block the farnesylation of CENP-E
RT and CENP-F and alter the association of CENP-E with the
RT microtubules.";
RL J. Biol. Chem. 275:30451-30457(2000).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: Belongs to the kinesin-like protein family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z15005; CAA78727.1; -
CC PIR; S28261; S28261.

HSP; P17119; 3KAR.
DR Genew; HGNC:1856; CENPE.
DR GK; Q02224; -
DR MIM; 117143; -
DR GO; GO:0005699; C:kinetochore; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0008350; F:kinetochore motor activity; TAS.
DR GO; GO:0000067; P:DNA replication and chromosome cycle; TAS.
DR GO; GO:0007079; P:mitotic chromosome movement; TAS.
DR GO; GO:0007080; P:mitotic metaphase plate congression; TAS.
DR InterPro; IPR001752; kinesin_motor.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS0067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere; Lipoprotein; Prenylation.
FT DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
FT LIPID 2660 2660 S-farnesyl cysteine.
SQ SEQUENCE 2663 AA; 312087 MW; CFCI3880C8C8B6 CRC64;
Alignment Scores:
Pred. No.: 2,25e-07 Length: 2663
Score: 266.50 Matches: 172
Percent Similarity: 38.30% Conservative: 126
Best Local Similarity: 22.11% Mismatches: 305
Query Match: 7.83% Indels: 175
DB: 1 Gaps: 30
US-09-602-362E-15 (1-2030) x CENE_HUMAN (1-2663)
Qy 12 AAAGATGCTCTTCTGAGGCTAACTCGGAATGAAATTTCTATTCCAACTAAGCCTTA 71
Db 1111 LysGluGlyLeuLeuSerArgThrCysAspArgLeuAlaGluValGluGluLys 1130
Qy 72 GAATTGATGACATGCAAACTTCAAGCAGAGCCTCCGAGAGGCATCTGCCTTCGAG 131
Db 1131 Glu-----LysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlu 1148
Qy 132 CTGCGCATTAAGAAATGCAAAAG-----TCTGTTCCAAATAAAGCCTTGGATTGAAG 182
Db 1149 GluMetSerGluMetGlnLysLysIleAsnGluIleGluAsnLeuLysAsnGluLys 1168
Qy 183 AAT---GAACAACATTTGAGCAGATGAGATATCTCCATCAGATCCAAACCAAAAG--- 236
Db 1169 AsnLysGluLeuThrLeuGluHisMetGluThrGluArgLeuGluAlaGlnLysLeu 1188
Qy 237 -----GACTATGAAGAAAGTCTTCTGGGATCTGTGAGAGTCTGTGAGACTGTTTCACAG 290
Db 1189 AsnGluAsnTyrGluGlu-----ValLysSerIle-----ThrLysGluArg 1202
Qy 291 AAGATGTGTGTTTACCAAGGCTACACATCAAAAGAAATAGATAAATAATGGAATA 350
Db 1203 LysValLeuLysGluLeuGlnLysSerPheGluThrGluArgAspHisLeuArgGlyTyr 1222
Qy 351 TTGAAGAGTCTCTCTGATTAATGATGTTTCTCGAAGCTCCCTGCAGATGAAAGTTTCT 410
Db 1223 IleArgGlu----- 1225
Qy 411 ATTCCAACCTAAGCCTTAGAATTGATGGACATGCACAACTTCAAGCAGAGCCTCCCGAG 470
Db 1226 -----IleGluAlaThrGlyLeuGlnThrLysGluLeuLysIleAlaHisIleHis 1243
Qy 471 AAGCCATCTGCCTCGAGCGCTGCATTGAATGAAAGTCAAAAGTCTGTTCCAAATAAAGCCTTG 530
Db 1244 LeuLysGluHisGlnGluThrIleAspGluLeuArgSerValSerGluLysThrAla 1263
Qy 531 GAATTGAAGATGAACAAACATTG----- 554

Db 1264 GlnIleAsnThrGlnAspLeuGluLysSerHisThrLysLeuGlnGluIlePro 1283
 QY 555 -----AGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAA 605
 Db 1284 ValLeuHisGluGlnGluLeuLeuProAsnValLysLysValSerGluThrGlnGlu 1303
 QY 606 AATCTTGGGATTCAGAGTCTCCGTGAGACTGTTTCAGAGGATGTGTGTACCC 665
 Db 1304 ThrMetAsnGluLeuGluLeuThrGluGlnSerThrLysAspSerThrThrLeu 1323
 QY 666 AAGGCTACACATCAAAAGAAATG-----GATAAATAAGTGGRAAATTAAGAGAT 716
 Db 1324 AlaArgIleGluMetGluArgLeuArgLeuAsnGluLysPheGlnGluSerGlnGluGlu 1343
 QY 717 TCACTAGCTATCAAAATCTTTGGATACAGTTCATTCTTGTGAAGAGCAAGGAACTT 776
 Db 1344 IleLysSerLeuThrLysGluArgAspAsnLeuLysThrIleLysGluAlaLeuGluVal 1363
 QY 777 CAA-----AAAGATCACTGTGAACACGTACAGGAAATGGAACAAATGAA 824
 Db 1364 LysHisAspGlnLeuLysGluHisIleArgGluThrLeuAlaLysIleGlnGluSerGln 1383
 QY 825 AAGAAG-----TTTGTGTACTGAAAAAGAACTGTCAAGACCAAAAGAAATAAAA 875
 Db 1384 SerLysGlnGluGlnSerLeuAsnMetLysGluLysAspAsnGluThrThrLysIleVal 1403
 QY 876 TCACGTTAGACACCAAAA-----GTTAATGGGAACAA 911
 Db 1404 SerGluMetGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIleGluMet 1423
 QY 912 GAGCTCTGAGTGTGAGTTGACTTTAAACCAAGAGAGAGAGAG----- 956
 Db 1424 LeuGlyLeuSerLysArgLeuGlnGluSerHisAspGluMetLysSerValAlaLysGlu 1443
 QY 957 -----AGAGAAATGCCGATATATTAATGAA 983
 Db 1444 LysAspLeuGlnArgLeuGlnGluValLeuGlnSerGluSerAspGlnLeuLysGlu 1463
 QY 984 AAAATTAGGGAAGATTAGGAG----- 1007
 Db 1464 AsnIleGluLeuValAlaLysHisLeuGluThrGluGluLeuLysValAlaHis 1483
 QY 1008 -----ATCCAGACAGCAGCAT-----AGGAAA 1028
 Db 1484 CysCysLeuLysGluGlnGluThrIleAsnGluLeuArgValAsnLeuSerGluLys 1503
 QY 1029 GAGTTAGAA-----GTGAACAACACTTGACAG-----GCTCTC 1064
 Db 1504 GluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLeuGlnAsn 1523
 QY 1065 AGAATACAGATATAGAAATGAAGAGTGTAGAAAGTAATTTGAATCAGGTTTCTCACACT 1124
 Db 1524 LysIleGlnGluIleLysGluLysGluGlnLeuAsnIleLysGlnIleSerGluVal 1543
 QY 1125 CATGAAATGAATAATATCTCTTACATGAAATTCGATGTGAAAAGCAATGCCATG 1184
 Db 1544 GlnGluAsnValAsn-----GluLeuLysGln 1552
 QY 1185 CTAAACTCGAATAGCCACACTGAACACCAATACAGGAAAGGAAATAAATATCTTT 1244
 Db 1553 PheLysGluHisArgLysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeu 1572
 QY 1245 GAGGACATTAAGATTTTAAAGAAAAGAAATGCTGAATTCAGATGACCTTAAACTGAAA 1304
 Db 1573 GluLeuThrAsnArgLeuGlnGluSerGlnGluGluIleGlnIleMetIleLysGluLys 1592
 QY 1305 GAGGAATCATTAACATAAGGGCCATCTCAATATAGTGGCGAGCTTAAGATTCATGAGCT 1364
 Db 1593 GluGlu-----MetLysArgValGlnGlnAlaLeuGlnIleGluArgAspGlnLeuLys 1610
 QY 1365 GAGAACACAAATG---CTCATTCTTAATTTGAGGAAAAACAGACAAAGAAATACTAGAG 1421
 Db 1611 GluAsnThrLysGluIleValAlaLysMetLysGluSerGlnGluLysGlu-Tyr----- 1628

QY 1422 GCAGAAATTTGAATCACACCATCTAGACTGGCTTCTGCTACAAAGACCATGATCAAAT 1481
 Db 1629 -GlnPheLeuLysMetThrAlaValAsn-----GluThrGlnGluLysMe 1643
 QY 1482 GTGACATCAAGAAAGAACTCAAGAACCTGCTTTCCACATTCAGAGAGATGCTTTTGGCAA 1541
 Db 1643 TCysGluIleGluHisLysGluGlnPheGluThrGlnLysLeuAsnLeuGluAsnIle 1663
 QY 1542 AGAAATTAATGATGTTGATGTGAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACC 1601
 Db 1663 eGlu-----ThrGluAsnIleArgLeuThrGlnIleLeuHisGluAs 1677
 QY 1602 ACTTTCTGAAGCTCAAGAGAAATCC-----ANAAAGCTCAAAAATAA 1643
 Db 1677 nLeuGluGluMetArgSerValThrLysGluArgAspAspLeuArgSerValGluGluTh 1697
 QY 1644 TCTCAATATTCAGAGAGATCTCTAAGAGAAATACATGCTTTCAGGACACATGCACAA 1703
 Db 1697 rLeuLysValGluArgAspGlnLeuLysIleValHisMetHisLysLysGluHisGlnG 1715
 QY 1704 GAGAC-----CAACGTGAACACACAG---TGTCAAATGAAGAGAGCTGAACACATGT 1751
 Db 1715 rGAspLeuGluLysGlnGluLeuLysIleValHisMetHisLysLysGluHisGlnG 1735
 QY 1752 NTCAAACGACAA-----CATNATGTGAACAAACACACATGCACAGC 1793
 Db 1735 luthrIleAspLysLeuArgGlyIleValSerGluLysThrAsnGluIleSerAsnMetG 1755
 QY 1794 AGGAGCTCTAGATCAGAAATATTTCACTCAAGCAAAATATGCTGGCTTCAACAGC 1853
 Db 1755 lnyAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspLeuLysIleGlnGluG 1775
 QY 1854 AATTA-----GTTATGACATTAAGAAGCTGACAAACAAGCAAGATAACAATTGAT 1907
 Db 1775 luLeuArgIleAlaHisMetHisLysLysGluGlnGlnGluThrIleAsp----- 1791
 QY 1908 NTTCAATNTCTTGAGAGAAATGCNATCATCTTCTTAAAGAGAAATAGGAGATAT 1967
 Db 1792 -----LysLeuArgGlyIleValSerGluLysThrAspLysLeuS 1805
 QY 1968 TTNATTACNATAACCATTTAAATAAACCGTATATTTCAATATGGAAAAA 2017
 Db 1805 erAsnMetGlnLysAspLeuGluAsnSerAsnAlaLysLeuGlnGluLys 1821

RESULT 9

ID NUFL_YEAST STANDARD; PRT; 944 AA.
 AC P32380;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE NUFL protein (Spindle poly body spacer protein SPC110).
 GN NUFL OR SPC110 OR YDR356W OR D9476.3.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8288C;
 RX MEDLINE=92176232; PubMed=1541631;
 RA Mirzayan C., Copeland C.S., Snyder M.;
 RT "The NUFL gene encodes an essential coiled-coil related protein that
 RT is a potential component of the yeast nucleoskeleton.";
 RL J. Cell Biol. 116:1319-1332(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94064779; PubMed=7503995;
 RA Kilmartin J.V., Dyos S.L., Kershaw D., Finch J.T.;
 RT "A spacer protein in the Saccharomyces cerevisiae spindle poly body
 RT whose transcript is cell cycle-regulated.";
 RL J. Cell Biol. 123:1175-1184(1993).

between the actin and microtubule cytoskeleton. May link endocytic vesicles to microtubules. May play a role in formation of furrows during cellularization.

!- SUBUNIT: Interacts with Lva.

!- SUBCELLULAR LOCATION: Cytoplasmic; microtubule-associated. Lva-CLIP-190 complexes are found at the Golgi.

!- ALTERNATIVE PRODUCTS:

Event=Alternative splicing; Named isoforms=3;

Name=A;

isoId=Q9VJB5-1; Sequences=Displayed;

Name=B;

isoId=Q9VJB5-2; Sequences=VSP_050479;

Note=No experimental confirmation available;

Name=C;

isoId=Q9VJB5-3; Sequences=VSP_050480;

Note=No experimental confirmation available;

!- TISSUE SPECIFICITY: Specifically expressed at the tip of the furrow in cellularizing blastoderms. CLIP-190 and jar are coexpressed at several times in development and in a number of tissues, including embryonic axonal neuron processes and posterior pole.

!- SIMILARITY: Contains 2 CAP-Gly domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL; AF041382; AAB96783.1; -

EMBL; AE003655; AAF53604.1; -

EMBL; AE003655; AAF53605.2; -

EMBL; AE003655; AAN10987.1; -

EMBL; AF118896; AAM50756.1; -

FLYBase; FBgn0020503; CLIP-190.

GO; GO:0005938; C:cell cortex; IDA.

GO; GO:0005794; C:Golgi apparatus; IDA.

GO; GO:0005875; C:microtubule associated complex; IDA.

GO; GO:0003779; F:actin binding; IDA.

GO; GO:0008017; F:microtubule binding; IDA.

GO; GO:0005515; F:protein binding; IMP.

GO; GO:0007349; P:cellularization; IMP.

InterPro; IPR000938; CAP-Gly.

Pfam; PF01302; CAP_GLY; 2.

PROSITE; PS00845; CAP_GLY_1; FALSE_NEG.

PROSITE; PS0245; CAP_GLY_2; 2.

KW Cytoskeleton; Golgi stack; Microtubule; Actin-binding; Coiled coil; Repeat; Alternative splicing.

KW Repeat; Alternative splicing.

FT DOMAIN 143 185 CAP-GLY 1.

FT DOMAIN 260 302 CAP-GLY 2.

FT DOMAIN 378 468 COILED COIL (POTENTIAL).

FT DOMAIN 484 660 COILED COIL (POTENTIAL).

FT DOMAIN 488 1452 GAU-RICH.

FT DOMAIN 667 946 COILED COIL (POTENTIAL).

FT DOMAIN 926 981 COILED COIL (POTENTIAL).

FT DOMAIN 1001 1121 COILED COIL (POTENTIAL).

FT DOMAIN 1158 1549 COILED COIL (POTENTIAL).

FT DOMAIN 1200 1461 GLN-RICH.

FT DOMAIN 1565 1600 COILED COIL (POTENTIAL).

FT VARSPIC 348 348 Missing (in isoform B).

FT /Ftrdvsp_050479.

FT VARSPIC 1 109 MSDDTSAGGTSAFPSPVTPADPEPCATASKLPGRINRNP

FT TPATSGTGPQSKMKAPSSFGSTGSKIGRPPCCNHTTPK

FT SGPPPEATSMGRSDNLSINSAYT -> MSRESDDNLS

FT SINSAYTDLYQETVRRFRSSLSPTPDWDRFSPARRSLKSE

FT AGRASYDYVLEATGRRRS (in isoform C).

FT /Ftrdvsp_050480.

FT S -> N (IN REF. 1).

FT D -> G (IN REF. 1).

FT K -> Q (IN REF. 1).

FT CONFLICT 207 207

FT CONFLICT 420 420

FT CONFLICT 492 492

FT CONFLICT 561 561 E -> A (IN REF. 1).

FT CONFLICT 614 614 T -> S (IN REF. 1).

FT CONFLICT 683 683 S -> I (IN REF. 1).

FT CONFLICT 692 692 N -> Q (IN REF. 1).

FT CONFLICT 717 717 M -> K (IN REF. 1).

FT CONFLICT 769 769 F -> L (IN REF. 1).

FT CONFLICT 787 787 Q -> E (IN REF. 1).

FT CONFLICT 881 881 Q -> E (IN REF. 1).

FT CONFLICT 907 909 HLL -> QLQ (IN REF. 1).

FT CONFLICT 920 920 G -> E (IN REF. 1).

FT CONFLICT 929 929 C -> Y (IN REF. 1).

SQ SEQUENCE 1690 AA; 189063 MW; D6F7916A9C532F16 CRC64;

Alignment Scores:

Pred. No.: 3 87e-07 Length: 1690

Score: 262.50 Matches: 170

Percent Similarity: 37.96% Conservative: 131

Best Local Similarity: 21.44% Mismatches: 263

Query Match: 7.71% Indels: 229

DB: 1 Gaps: 34

US-09-602-362E-15 (1-2030) x CI90_DROME (1-1690)

Qy 39 GGAATGAAGTTCTATTCCAACTAAAGCCTTAGAATTGATGACATG-----CAAACT 92

Dy 330 GlyThrMetAsnSerIleAlaThrAlaThrSerArgMetArgMetAsnAlaGln 349

Qy 93 TTCAAGCAGACGCTCCGAGAACCATCTGCTTCGAGCCTGCCATTGAAATG-----146

Dy 350 ArgLysSerSerThrProValLysProIleLeuAlaThrProLysSerGlnPheSerMet 369

Qy 146 -----146

Dy 370 GlnAspLeuLeuArgGluLysGlnGlnHisValGluLysLeuMetValGluArgAspLeu 389

Qy 147 ---CAAAGTCTGTTCCAAATAAAGCCTTGGAATTG---AAGATGAACAAACATGAGA 200

Dy 390 AspArgGluAspAlaGlnAsnGlnAlaLeuGlnLysAsnIleAsnGluLeuLys 409

Qy 201 GCA-----GATGAGATACCTCCATCAGAAATCCAAACAAAGAGACTATGAA 245

Dy 410 AlaArgIleValGluLeuGluSerAlaLeuAspAsnGluArgLysLysThrGluGluLeu 429

Qy 246 GAAGTCTTGGGATCTGAGACTCTGT-----GAGACTTTTCACAGAGAGATGTG 299

Dy 430 GlnCysSerIleAspGluAlaGlnPheCysGlyAspGluLeuAsnAlaGlnSerGlnVal 449

Qy 300 TGTTTACCCAGGCTACACATCAAAAGAAATAGATAAAATAAATGAAAAATTA-----353

Dy 450 -----TyrLysGluLysIleHisAspLeuGluSerLysIleThrLys 463

Qy 354 -----GAAGAGTCTCTGTATATGAT 374

Dy 464 LeuValSerAlaThrProSerLeuGlnSerIleLeuProAspLeuProSerAspAsp 493

Qy 375 GGTCTTCTGAAG-----GCTCCCTGCAGAAATGAAGTTCTTATTCACACTAAAGCC 425

Dy 484 GlyAlaLeuGlnGluIleAlaLysLeuGlnLysMetThrIleGlnGlnLysGlu 503

Qy 426 TTAGAATTGATG-----437

Dy 504 ValGluSerArgIleAlaGluGlnLeuGluGlnArgLeuArgGluAsnValLys 523

Qy 438 -----GNATGCAAACTTTCAAGCAGAGCCTCCCGAGAGCCATCTGCCTTC 485

Dy 524 TyrLeuAsnGluGlnIleAlaThrLeuGlnSerGlnLeuValSerLysAspGluAlaLeu 543

Qy 486 GAGCCTGCCATTGAAATGCAAAAGTCTGTTCCCAAT-----AAAGCCTTGGAAATGAAG 539

Dy 544 GluLysPheSerLeuSerGluCysGlyIleGluAsnLeuArgGluLeuGluLeuLeu 563

Qy 540 AATGAACAACATGTGAGAGAGATGATGTTCCCTTCAGAAATCAAAACAAAGAGAGTT 599

Db 564 LysGluGluAsnGluLysGlnAlaGlnAlaGluPheThrArgLysLeuAla 583
 QY 600 GAAGAAAT-----TCTGGGATTC 620
 Db 584 GluLysSerValGluValLeuArgLeuSerSerGluLeuGlnAsnLeuLysAlaThrSer 603
 QY 621 GAGAGTCTCCGT---GAGACTGTTTACAGAGGAGTGCTGTGTACCCAGGCTACACAT 677
 Db 604 AspSerLeuGluSerGluArgValAsnLysThrAspGluCysGluLeuLeuGlnThrGlu 623
 QY 678 -----CAAAAGCAATGATTAATAAGTGA 704
 Db 624 ValArgMetArgAspGluGlnIleArgGluLeuAsnGlnLeuAspGluValThrThr 643
 QY 705 AAATTAGAA-----GATTCAACTAGCTATCAAAATCTTGCATACAGTTTAT 752
 Db 644 GlnLeuAsnValGlnLysAlaAspSerSerAlaLeuAspMetLeu----- 659
 QY 753 TCTGTGAAGACAGGAACTTCAAAAGATCACTGTGAACACGTCACAGGAAAATG 812
 Db 660 -----ArgLeuGlnLysGluGlyThrGluGluLysSerThrLeuLeu 673
 QY 813 GAACAAATGAAGAAGTTTCTGTACTCAAAAGAACTGTCAGACCAAGAAATA 872
 Db 674 GluLys-----ThrGluLysGluLeuValGlnSerLysGluGln 686
 QY 873 AAATCACTAGTAGAGAACCAAAAGTTAAATGGGAACAAGAGCTCGAGTGTGAGATTG 932
 Db 687 AlaAlaLysThrLeuAsnAspLysGluGlnLeuGluLysGlnIleSerAspLeuLysGln 706
 QY 933 ACTTTAAACCAAGACAGAGACAGAAATGCCGATATATTAATCAAAATATTAG 992
 Db 707 LeuAlaGluGlnLysLeuValArgGluMetThrGluAsnAlaIleAsnGlnIleGln 726
 QY 993 GAAGAAATTAGCAAGATTCGAAGACAG-----CATAGAAAGAGTTAGAA----- 1037
 Db 727 LeuGluLysGluSerIleGluGlnGlnLeuAlaLeuLysGlnAsnGlnLeuGluAspPhe 746
 QY 1038 GTCAAAACAACACTTGAACAGGCTCTCAGATACAGATATAGAA----- 1082
 Db 747 GlnLysLysGlnSerGluValHisLeuGlnGluIleLysAlaGlnAsnThrGln 766
 QY 1083 -----TTGAAGAGTGTAGAAAGTAAATTGAT 1109
 Db 767 LysAspPheGluLeuValGluSerGlyGluSerLeuLysLysGlnGlnGlnLeuGlu 786
 QY 1110 CAGGTTTCTCACATCATGAAATGAAATATCTCTTACATGAAATTCGATGTGAA 1169
 Db 787 GlnLysThrLeuGlyHisGluLysLeuGlnAlaLeuGluGlu-----LeuLys 803
 QY 1170 ARGGA-----ATTGCCATGCTAAACTGGAATAGCCACACTGAACACCAA 1217
 Db 804 LysGluLysGluThrIleLysGluLysGluGlnGluLeuGlnGlnLeuGlnSerLys 823
 QY 1218 TACAGGAAAGGAAAT-----AAATCTTTAGACACTT 1253
 Db 824 SerAlaGluSerGluSerAlaLeuLysValGlnValGlnLeuGlnLeuGlnGln 843
 QY 1236 -----AAATCTTTAGACACTT 1253
 Db 844 GlnAlaAlaSerGlyGluGlySerLysThrValAlaLysLeuHisAspGluIle 863
 QY 1254 AAGATTTTAAAGAAAGAAATGCTGAATTCAGATGACCTTAACTGAAAGAGAAATCA 1313
 Db 864 SerGlnLysSerGlnAlaGluGluThrGlnSerGluLysSerThrGlnSerAsn 883
 QY 1314 TTAACATAAAGGCATCTCAAT-----AGTGGCAGCTTAAAGTTCTGATAGCT 1364
 Db 884 LeuGluAlaLysSerLysGlnLeuGluAlaAlaAsnGlySerLeuGlu----- 899
 QY 1365 GAGACACAAATGCTCACTTCTAAATTAAGGAAAAACACACAAAGAAATACCTAGAGCA 1424
 Db 900 GluGluAlaLysLysSerGlyHisLeuLeuGluGlnIleThrLys-----LeuLysSer 917

QY 1425 GAAATTTGAATCACACCACCTCTAGACTCGCTTCTGCTCTCAAGACCATGATCAATTTGTG 1484
 Db 918 GluValGlyGluThrGlnAlaAlaLeuSerSerCys-----HisThrAspValGlu 934
 QY 1485 ACATCAAGAAAGAACTCAAGAACCTGCTTTCCACATTCACAGAGATGCTTTTTCGAAAGA 1544
 Db 935 SerLysThrLysGlnLeuGlu-----AlaAlaAsnAlaAlaLeuGlu----- 948
 QY 1545 AAAATGATGTTGAT-----GTGAGTAGTACCGATATATATCAATGAGGT 1589
 Db 949 LysValAsnLysGluThrAlaGluSerArgAlaGluAlaSerAspLeuGlnAspLysVal 968
 QY 1590 G-----CTCCATCAACACCTTCTTGAAGCTCAAGAGGAAATCCANAAG 1631
 Db 968 LysGluIleThrAspThrLeuHisAlaGluLeuGlnAlaGluArgSerSerSerAl 988
 QY 1632 CCTAAAATTAATCTCAATTATCGAGAGATGCTCTTAAGAGAAATACAT----- 1681
 Db 988 aLeuHisThrLysLeuSerLysPheSerAspGluIleAlaThrGlyHisLysGluLeu 1008
 QY 1682 -----TGGTTTCAGGAACATGCACAAAGACACCAACGTCGAAACACAGT 1724
 Db 1008 hrSerLysAlaAspAlaTrpSerGlnGluMetLeuGlnLysGluLys----- 1023
 QY 1725 GTCAAAATGAAGGAGCTGAACACATGNTCAANCGAACCAAGATNATGTGAACAAACACA 1784
 Db 1024 --GluGlnGlnLeuArgGlnGlnLeuGlnAspSerGlnAspSerGlnThrLysLeu 1043
 QY 1785 CTGANCAAGAGGAGCTCTAGATCAGAAATTTTCAACTACAAAGCAAAATATGTGGC 1844
 Db 1043 ysAlaGluGlyGluArg---LysGluLysSerPheGluGluSerIleLysAsn-----L 1060
 QY 1845 TTCACACGCAATAGTTTCATGCACATAAAGAAAGC 1879
 Db 1060 euGlnGluGluValThrLysAlaLysThrGluAsn 1071
 RESULT 11
 ID MYH9_RAT STANDARD; PRT; 1961 AA.
 AC Q62812;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
 DE type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
 GN MYH9.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Choi O.H., Park C., Itoh K., Adelstein R.S., Beaven M.A.;
 RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Cellular myosin appears to play a role in cytokinesis,
 CC cell shape, and specialized functions such as secretion and
 CC capping.
 CC -!- SUBUNIT: Myosin is a hexameric protein that consists of 2 heavy
 CC chain subunits (MHC), 2 alkali light chain subunits (MLC) and 2
 CC regulatory light chain subunits (MRC-2).
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)


```
CC or send an email to license@iebs-sib.ch).
CC -----
DR EMBL; U01463; AAA74950.1; -.
DR HSP; P10587; 1BR2.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; Myosin head.
DR InterPro; IPR004009; Myosin N.
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00966; IQ; 1.
KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
KW Coiled coil; Multigene family.
FT DOMAIN 1 778 MYOSIN HEAD-LIKE.
FT DOMAIN 779 808 IQ.
FT DOMAIN 841 1927 COILED COIL (POTENTIAL).
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 654 676 ACTIN-BINDING.
SQ SEQUENCE 1961 AA; 226336 MW; 959876D9681FB19E CRC64;

Alignment Scores:
P-red. No.: 4.03e-07 Length: 1961
Score: 262.00 Matches: 164
Percent Similarity: 38.44% Conservative: 137
Best Local Similarity: 20.95% Mismatches: 284
Query Match: 7.70% Indels: 198
DB: 1 Gaps: 30

US-09-602-362E-15 (1-2030) x MYH9_RAT (1-1961)
QY 141 GAAATGCAAAAGTCTGTTCCAAATAAAGCTTGGAAATG---AAGATGAACAACAATG 197
D 1066 GluLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGluGluLeu 1085
QY 198 ACAGCA-----GATGAGATACTCCCATCAGATCCCAACAAAGGAC 239
D 1086 GlnAlaAlaLeuAlaArgValGluGluAlaAlaGlnLysAsnMetAlaLeuLys 1105
QY 240 TATGAAGAAAGTCTCTGGGATCTCAGAGTCTCTGTGACGTCTTACAGAGATG 299
D 1106 IleArgGluLeuGluThrGlnIleSerGluLeuGlnGluAspLeuGluSerGluArgAla 1125
QY 300 TCTTTACCAAGCTACACATCAAAA-----GAATAGATAAATAAT 344
D 1126 CysArgAsnLysAlaGluLysGlnLysArgAspLeuGluGluLeuAlaLeuLys 1145
QY 345 GGAATAATTAGAGAGTCTCTGATATGATGTTT-----CTGAAGCTCCCTGC 395
D 1146 ThrGluLeuGluAspThrLeuAspSerThrAlaAlaGlnGlnGluArgSerLysArg 1165
QY 396 AGAATGAAGATTTCTATTCCTCAACTAAGCTTAGATTTGATGGACATCAACATTTCAA 455
D 1166 GluGlnGluValSerIleLeuLysLysThrLeuGlu---AspGluAlaLysThrHisGlu 1184
QY 456 GCAGAGCTCCCGAAGAGCCATCTGCTTCGAGCTCGCATTTGAA----- 500
D 1185 AlaGlnIleGlnGluMetArgGlnLysHisSerGlnAlaValGluGluLeuAlaGluGln 1204
QY 501 -----ATGCAAAAGTCTGTTCCCAAT 521
D 1205 LeuGluGlnThrLysArgValLysAlaThrLeuGluLysAlaLysGlnThrLeuGluAsn 1224
QY 522 AAAGCTTCGAATTGAAGATGAACAAACATTCAGAGCAGATCAGATGTTCCCTTCAGAA 581
D 1225 GluArgGluGluLeuAlaAsnGluValLysAlaLeuLeuGlnGlnLysGlyAspSerGlu 1244
QY 582 TCAAAACAAAGAGGTTGAAGAAAATTTCTGG-----GATTCT 620
```

```
1245 HisLysArgLysLysValGluAlaGlnLeuGlnValLysPheSerGluGly 1264
QY 621 GAGAGTCTCGTGAGACTGTTTCCACAGAGGATGTGTGTACCCAGGCTACACATCAA 680
D 1265 GluArgValArgThrGluLeuAlaAspLys-----ValSerLysLeu-----Gln 1279
QY 681 AAAGAAATGGATAAATAAGTGA----- 704
D 1280 ValGluLeuAspSerValThrGlyLeuLeuAsnGlnSerAspSerLysSerLysLeu 1299
QY 704 ----- 704
D 1300 ThrLysAspPheSerAlaLeuGluSerGlnLeuGlnAspThrGlnGluLeuGlnGlu 1319
QY 705 -----AAATTAGAGATTCAACTAGCCTATCAAAATCTTGATACAGTTTCAT 752
D 1320 GluAsnArgGlnLysLeuSerLysLeuSerThrLysLeuLysGlnMetGluAspGluLysAsn 1339
QY 753 TCTTGTGAAGAGCAAGGAACTTCAAAAAGATCACTGTGAACAACGTCACAGAAA--- 809
D 1340 SerPheArgGluGlnLeuGluGluGluGluAlaLysArgAsnLeuGluLysGln 1359
QY 810 -----ATGGAACAAATGAAAAAGATTT----- 833
D 1360 IleAlaThrLeuHisAlaGlnValThrAspMetLysLysMetGluAspGlyValGly 1379
QY 834 TGTGTA-----CTGAAA 845
D 1380 CysLeuGluThrAlaGluGluAlaLysArgArgLeuGlnLysAspLeuGluGlyLeuSer 1399
QY 846 AAGAACTGTGAGAACAAAGAAATAAATCATCAGTTAGAGAACCAAAAGTTAAATGG 905
D 1400 GlnArgLeuGluGluLysValAlaAlaTyrAspLysLeuGluLysThrLysThrArgLeu 1419
QY 906 GAACAAGAGCTCTGCAGTGTGAGATTCAGTTTAAACCAAGAA----- 947
D 1420 GlnGlnLeuLeuAspLeuValAspLeuValAspHisGlnArgGlnSerValSerAsn 1439
QY 948 ---GAAGAGAGAGAGAAATGCCGAT-----ATATTAAATGAA 983
D 1440 LeuGluLysLysGlnLysLysPheAspGlnLeuLeuAlaGluGluLysThrIleSerAla 1459
QY 984 AAAATTAGGAAGATTTAGGAAGATCGA-----CAGCAGCATAGGAAGAG 1031
D 1460 LysTyrAlaGluGluArgAspArgAlaGluAlaGluAlaArgGluLysGluThrLysAla 1479
QY 1032 TTAGAGTGAACAACAACCTTGAACAGGCTCTCAGAAATCAAGATATAGAAATG--- 1085
D 1480 LeuSerLeuAlaArgAlaLeuGluGluAlaMetGlu---GlnLysAlaGluLeuGluArg 1498
QY 1086 -----AAGAGTGTAGAAATTAATTCAGTTTCTCAGACTCATGAAATGAAAT 1139
D 1499 LeuAsnLysGlnPheArgThrGluMetGluAspLeuMetSerSerLysAspValGly 1518
QY 1140 TATCTCTTACATCAAAATTCATGTTGAAAAGGAAATTCGCCATGCTAAACCTGGAATA 1199
D 1519 LysSerValHisGlu-----LeuGluLysSerAsnArgAlaLeuGluGlnGlnVal 1535
QY 1200 GCCACACTGAACACCAATACACAGGAAAAGGAAAATAATATAC-----TTTGAGGACATT 1253
D 1536 GluGluMetLysThrGlnLeuGluGluLeuGluAspGluLeuGlnAlaThrGluAspAla 1555
QY 1254 AAGATTTTAAAGAAAAGAAAT-----GCTGAACCTTCAGATGACCCATAAAA 1298
D 1556 LysLeuArgLeuGluValAsnLeuGlnAlaMetLysAlaGlnPheGluArgAspLeuGln 1575
QY 1299 CTGAAGAGGAATCATTAACATAAAGGGCATCTCAATATATAGTGGCAGCTTAAAGTTCTG 1358
D 1576 GlyArgAspGluGlnSerGluGluLysLysGlnLeuValArgGlnValArgGluMet 1595
QY 1359 ATAGCTGAG-----AACACATGCTCCTCTTAATTTGAAG 1394
```

Db 1596 GluAlaGluLeuGluAspGluArgLysGlnArgSerIleAlaMetAlaAlaArgLysLys 1615
 Qy 1395 GAAAAACAGACAGAAAGAAATATAGAGGAGGAGAAATTAATATCATCATCTAGACTGGCT 1454
 Db 1616 LeuGluMetAspLeuLysAspLeuGluAlaHisIleAspThrAlaAsnLysAsnArgGlu 1635
 Qy 1455 TCTGCTGTACACACCATGATCAATTTGTGCATCAAGAAAAAGTCAGAACTCTGTTTC 1514
 Db 1636 GluAlaIle-LysGlnLeuArgLysLysLysLysLysLysLysLysLysLysLysLys 1653
 Qy 1515 CATCTTCAGGAGATGCTGTTTTCAGAAAGAAATGAATGTTGATGTGTAGTAGTACCGAT 1574
 Db 1653 GAspValAspThrArgAlaSerArgGluLysLysLysLysLysLysLysLysLysLys 1663
 Qy 1575 ATATAAATGAGTGTCTCATCAACCACTTCTTGAAGCTCAAGGAGATCCANAAAGCT 1634
 Db 1664 -----GluIleLeu-----AlaGlnAlaLysGluAsnGluLys 1674
 Qy 1635 AAAAATTAATCTCAATTTATGAGGAGATGCTCTTAAGAGAAATATACATTTGTTTCAAGAAC 1694
 Db 1674 sLysLeuLys-SerMetGluAlaGluMetIle-----GlnLeuGlnGluG 1689
 Qy 1695 AT-----GCACAAGAGACCACTGAAACACACAGTGTCAATGAAGGAGCTGAAC 1745
 Db 1689 lueuAlaAlaAlaGluArgAlaLysArgGlnAlaGln-----GlnGluArgAspG 1706
 Qy 1746 ACATGNTTCAAAACGAACAGATNATGTGAACAAACACACACTGACAGGAGTCT----- 1801
 Db 1706 lueuAlaAspGluIleAlaAsnSerSerClyLysGlyAlaLeuAlaLeuGluLysA 1726
 Qy 1802 -----CTAGATCAGAAATATTTCACACTACAAAGCAAA 1835
 Db 1726 rgArgLeuGluAlaLeuIleAlaLeuLysGluGluGluGluGluGluGluGluGluG 1746
 Qy 1836 ATATGCTGCTCAACAGCATAGTTCATGCACATAAAGAAAGCTGACACAAAGCAAG 1895
 Db 1746 hrGluLeuIleAsnAspArgLeuLysAlaAsnLeuGlnIleAspGlnIleAsnThrA 1766
 Qy 1896 ATAAACAT-----TGATTTTCATTTCTGAGAGGAAATGCNATCATCTCTTCAAAG 1949
 Db 1766 spLeuAsnLeuGluArgSerHisAlaGlnLysAsnGluAsnAlaArgGlnLeuGluA 1786
 Qy 1950 AGAAATGAGGAGATTTTATACATACATTTTAAACACCGTATATTTCAATAT 2009
 Db 1786 rgGlnAsnLysGluLeuLysAlaLysLeuGlnGluMetGluSerAlaValLysSerLys 1806
 Qy 2010 GGAAG 2014
 Db 1806 yLys 1807

RESULT 12

MYHB_RABIT STANDARD; PRT; 1972 AA.
 AC P35748;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, smooth muscle isoform (SMWHC).
 GN MYH11.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92073350; PubMed=1961735;
 RA Babij P., Kelly C., Periasamy M.;
 RT "Characterization of a mammalian smooth muscle myosin heavy-chain
 RT gene: complete nucleotide and protein coding sequence and analysis of
 RT the 5' end of the gene";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10676-10680 (1991).
 CC -!- FUNCTION: Muscle contraction.
 CC -!- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2

CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -!- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -!- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -!- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -!- SIMILARITY: Contains 1 IQ domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; M77812; AAA31395.1; -.
 HSSP; P10587; 1BR2
 InterPro; IPR000048; IQ region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR002017; Spectrin.
 Pfam; PF00612; IQ; 1.
 Pfam; PF00063; myosin_head; 1.
 Pfam; PF02736; myosin_N; 1.
 Pfam; PF01576; Myosin_tail; 1.
 PRINTS; PR00193; MYOSINHEAVY.
 ProDom; PD000355; myosin_head; 1.
 SMART; SMC0015; IQ; 1.
 SMART; SMC0242; MYSC; 1.
 PROSITE; PS00096; IQ; 1.
 Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 Calmodulin-binding; ATP-binding; Methylation; Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 807 IQ.
 FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
 FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 763 777 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 SQ SEQUENCE 1972 AA; 227318 MW; 2061A224288D6A4C CRC64;

Alignment Scores:

Pred. No.: 4,28e-07 Length: 1972
 Score: 261.50 Matches: 128
 Percent Similarity: 40.00% Conservative: 118
 Best Local Similarity: 20.81% Mismatches: 215
 Query Match: 7.68% Indels: 154
 DB: 1 Gaps: 18
 US-09-602-362E-15 (1-2030) x MYHB_RABIT (1-1972)

Qy 528 TTGAATGAGAGATGACAA----- 548
 Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846
 Qy 549 ---ACATTGAGAGCAGATCATGATGTTCCCTTCAGAAATCAAAACAAAGAGAGTTGAGAA 605
 Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysLysLysGlu 866
 Qy 606 AATTCTTGGAGTCTGAGACTCCCGTGAGACTGTTTCACAGAGAGATGTTGTGTACCC 665
 Db 867 ArgGlnGlnLysAlaGluSerGluLeuGlnGluGlnLysHisThrGlnLeu--- 885
 Qy 666 AAGCTACATCAACAAAGAAATGGAATAAATAGTGGAAATATTAGAGATTCAACTAGC 725

886 Db -----SerGluGluLysAsnLeuLeuGlnGluLeuGlnAlaGluThrGlu 901
726 QY CTA-----TCRAAAATCTTG 740
902 Db LeuTyrAlaGluAlaGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGlu 921
741 QY GATACAGTTCATCTTGTGAAGACAGCAAGGAACTTCAAAAAGATCCTGTGAAACAAGT 800
922 Db GGuileLeuHisGluMetGluAlaArgLeuGluGluGluGluGluGluGluGluGlu 941
801 QY ACAGGAAATGAAACAAATGAAAGAGATTGTTGTACTGAAAGAAAGAACTGTGAGAA 860
942 Db GlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGluGlnLeuGluGlu 961
861 QY GCAAGAGAAATATAATACAGTTA----- 884
962 Db GluGluAlaAlaArgGlnLysLeuGlnGluLysValThrAlaGluAlaLysLys 981
885 QY -----GAGACCAAAAAGTTAAATGGGAACA 911
982 Db LysLeuGluAspLysLeuValMetAspAspGlnAsnLysLeuSerLysLys 1001
912 QY GAGCTCTGAGTGTGAGATTGACT-----TTAAACCAAGAAAGAGAG 956
1002 Db LysLeuGluGluArgLysLeuThrThrAsnLeuAlaGluGluGluGluGlu 1021
957 QY AGAAGAAATGCCATATATAATGAAGAAATAGGAAGAAATAGGAAGATCGAAG 1016
1022 Db AlaLysAsnLeuThrLysLysLysLysLysHisGluSerMetLysLeuGluVal 1041
1017 QY CAGCATGAGGAAGAGTTAGAACTTGAACCAACCTTGAACGGCTCTCAGAAATACAAG 1076
1042 Db ArgLeuLysLysGluGluLysSerArgGlnGluLeuGluLysLysLysMetAsp 1061
1077 QY ATAGAATTGAAGAGTGAAGAGTAATTTGAATCAGGTTTTCACACTCATGAAATGAA 1136
1062 Db GlyGluAlaSerAsp----- 1066
1137 QY AATTATCTTCAATCAAAATTCATGTTGAAAAGAAATGCCATGCTAAACATGGA 1196
1067 Db -----LeuHisGlu-----GlnleAlaAspLeuGlnAlaGln 1077
1197 QY ATAGCCACATGAAACACCAATACCAGGAAAGAAATATAATCTTTGAGACATTAAG 1256
1078 Db IleAlaGluLysMetGlnLeuAlaLysLysLysGluGluGluAlaAlaLeuAla 1097
1257 QY ATTTAAAGAAAGAAAGTGAAGTCACTTCAATGACCTTAA----- 1298
1098 Db ArgLeuGluAspGluThrSerGlnLysAsnAsnAlaLeuLysLysLysLysLys 1117
1299 QY -----CTGAAAGAGAAATCATTAATCTAAAGGGCATCTCAATATAGTGG 1343
1118 Db GlyHisLysAspLeuGlnGluAspLeuAspSerGluArgAlaAlaArgAsnLysAla 1137
1344 QY -----CAGCTTAAAGTTCTCATAGCTGAG----- 1367
1138 Db GluLysGlnLysArgAspLeuGluGluLeuGluAlaLeuLysThrGluLeuGluAsp 1157
1368 QY -----AACCAATGCTCACTTCTAAATTTGAAGGAAACCAAGCAAGAAATA--- 1415
1158 Db ThrLeuAspThrThrAlaThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnThr 1177
1416 QY -----CTAGAGCAGAAATTTGAATCACCAATCTCAGCTGCTTCTGCT 1460
1178 Db ValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGluMet 1197
1461 QY GTACAGACCATGATCAAAATTTGACATCAAGAAAGTCAAGAACCTGCTTCCACATT 1520
1198 Db ArgGlnLysHisThrGlnValGluGluLeuLeuThrGluGlnGluGlnPheLysArg 1217
1521 QY GCAGGAGATGCTTGTGTTGCAAGAAAGAAATCAATGTTGATGTCAGTAGTACCATATATAA 1580
1218 Db AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeu-Al 1237

1581 QY CAATGAG-----GTGCTCCATCAACCACTTCTTGAAGCTCAAGAGAAATCCAAAGCCT 1634
1237 Db aGlyLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLe 1257
1635 QY AAAAATTAATCTC-----AATTATGAGGAGATGCTCTAAGAGAAATACATT 1682
1257 Db uGluValGlnLeuGlnGlnLysSerLysCysSerAspGlyGluArgAlaGln 1277
1683 QY GTTTTCAGGAACATGTCACAAAGAGACCAACGCTGAAACACACAGTGT-----CAATGA 1733
1277 Db uLeuAsn-AspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeu 1297
1734 QY AGGAAGCTGAA-----CACATGT 1751
1297 Db erGluAlaGluGlyLysAlaLysLeuAlaLysGluValAlaSerLeuGlySerGln 1317
1752 QY NTCAAAACGACCAAGATGTTGAAACAAACACACTGANCAGCAGGAGTCTAGATCAGA 1811
1317 Db euGlnAspThrGlnGluLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnValSerThrL 1337
1812 QY AATTATTTCAACTACAAAGCAAAATATGTGGTTTCAACAGCAATTAGTTTCATGCACATA 1871
1337 Db ysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnGlnGlnGlnGln-----A 1353
1872 QY ANGAAGCTGACACAAACAGACATCAACAATTCATTTTCTTCTTGAGAGAAATG 1931
1353 Db spGluGluMetGluAlaLysGln-----Asn-LeuGluArgHisLysSerThrLeuAsnile 1371
1932 QY CNATCATCTTCTTAAAGAGAAATGAGGAGATATTATTTATACNATPAACCATTTAAAAA 1991
1372 Db GlnLeuSerAspSerLysLysLysLeuGlnAspPheAlaSerThrVal----- 1387
1992 QY ACCGTATATTTCATATGCAAAAAAANAANAANAANAANAANAANAANAANAANA 2030
1388 Db ---GluSerLeuGluGluGlyLysLysLysLysLysLysLysLysLys 1399

RESULT 13
BEAL_HUMAN
ID BEAL_HUMAN STANDARD; PRT; 1411 AA.
AC Q15075; Q14221;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Early endosome antigen 1 (Endosome-associated protein p162).
GN BEAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=95286647; PubMed=7768953;
RA Mu F.-T., Callaghan J.M., Steele-Mortimer O., Stenmark H.,
RA Parton R.G., Campbell P.L., McCluskey J., Yeo J.-P., Tock E.P.C.,
RA Toh B.-H.;
RT "BEAL, an early endosome-associated protein. BEAL is a conserved
RT alpha-helical peripheral membrane protein flanked by cysteine
RT 'fingers' and contains a calmodulin-binding IQ motif.";
RL J. Biol. Chem. 270:13503-13511(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Seelig H.P.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP INTERACTION WITH RAB5A
RX MEDLINE=98361239; PubMed=9697774;
RA Simonsen A., Lippe R., Christoforidis S., Gaullier J.-M., Brech A.,
RA Callaghan J.M., Toh B.-H., Murphy C., Zerial M., Stenmark H.;
RT "BEAL links PI(3)K function to Rab5 regulation of endosome fusion.";
RL Nature 394:494-498(1998).
RN [4]

INTERACTION WITH RAB5A AND RAB5B.
RX MEDLINE=99421752; PubMed=10491193;
RA Callaghan J.M., Nixon S., Bucci C., Toh B.-H., Stenmark H.;
RT "Direct interaction of EEA1 with Rab5b.";
RL Eur. J. Biochem. 265:361-366(1999).
[5]
RX INTERACTION WITH STX6, AND SUBCELLULAR LOCATION.
RX MEDLINE=99436077; PubMed=10506127;
RA Simonsen A., Gaullier J.-M., D'Arrigo A., Stenmark H.;
RT "The Rab5 effector EEA1 interacts directly with syntaxin-6.";
RL J. Biol. Chem. 274:28857-28860(1999).
[6]
RX MUTAGENESIS OF ASP-1352; ASN-1357; 1367-VAL-THR-1368; ARG-1375 AND
RX ARG-1400, HOMODIMERIZATION, AND INTERACTION WITH PHOSPHATIDYLINOSITOL
RX 3-PHOSPHATE.
RX MEDLINE=99322673; PubMed=10394369;
RA Kutateladze T.G., Ogburn K.D., Watson W.T., de Beer T., Emr S.D.,
RA Burd C.G., Overduin M.;
RT "Phosphatidylinositol 3-phosphate recognition by the FYVE domain.";
RL Mol. Cell 3:805-811(1999).
[7]
RX MUTAGENESIS OF TRP-1349; CYS-1358; PHE-1365; ARG-1370; ARG-1371;
RX HIS-1372; HIS-1373; CYS-1374; ARG-1375; CYS-1377; GLY-1378; CYS-1385;
RX ARG-1400 AND CYS-1405, SUBCELLULAR LOCATION, AND INTERACTION WITH
RX PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=20387352; PubMed=10807926;
RA Gaullier J.-M., Roennig E., Gillooly D.J., Stenmark H.;
RT "Interaction of the EEA1 FYVE finger with phosphatidylinositol
RT 3-phosphate and early endosomes. Role of conserved residues.";
RL J. Biol. Chem. 275:24595-24600(2000).
[8]
RX INTERACTION WITH RAB22A.
RX MEDLINE=21859373; PubMed=11870209;
RA Kauppi M., Simonsen A., Bremnes B., Vieira A., Callaghan J.M.,
RA Stenmark H., Olkkonen V.M.;
RT "The small GTPase Rab22 interacts with EEA1 and controls endosomal
RT membrane trafficking.";
RL J. Cell Sci. 115:899-911(2002).
[9]
RX MUTAGENESIS OF GLU-39; PHE-41; ILE-42; PRO-44; MET-47 AND TYR-60,
RX HOMODIMERIZATION, AND INTERACTION WITH RAB5C.
RX MEDLINE=22499593; PubMed=12493736;
RA Merithew E., Stone C., Bathiraj S., Lambright D.G.;
RT "Determinants of Rab5 interaction with the N terminus of early
RT endosome antigen 1.";
RL J. Biol. Chem. 278:8494-8500(2003).
[10]
RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1289-1411 IN COMPLEX WITH
RX PHOSPHATIDYLINOSITOL 3-PHOSPHATE, AND HOMODIMERIZATION.
RX MEDLINE=21617582; PubMed=11741531;
RA Dumas J.J., Merithew E., Sudharshan E., Rajamani D., Hayes S.,
RA Lawe D., Corvera S., Lambright D.G.;
RT "Multivalent endosome targeting by homodimeric EEA1.";
RL Mol. Cell 8:947-958(2001).
[11]
RX STRUCTURE BY NMR OF 1346-1410 ALONE AND IN COMPLEX WITH
RX PHOSPHATIDYLINOSITOL 3-PHOSPHATE.
RX MEDLINE=21143489; PubMed=11230696;
RA Kutateladze T.G., Overduin M.;
RT "Structural mechanism of endosome docking by the FYVE domain.";
RL Science 291:1793-1796(2001).
CC -!- FUNCTION: Binds phospholipid vesicles containing
CC phosphatidylinositol 3-phosphate and participates in endosomal
CC trafficking.
CC -!- SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC RAB22A that have been activated by GTP-binding.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and peripheral membrane protein
CC associated with early endosomes.
CC -!- DOMAIN: The FYVE-type zinc finger domain mediates interactions
CC with phosphatidylinositol 3-phosphate.
CC -!- DISEASE: Antibodies against EEA1 are found in sera from patients
CC with subacute cutaneous lupus erythematosus and other autoimmune
CC diseases.

CC -!- SIMILARITY: Contains 1 C2H2-type zinc finger.
CC -!- SIMILARITY: Contains 1 FYVE-type zinc finger.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L40157; AAA79121.1; -;
DR EMBL; X78998; CAA55632.1; -;
DR PIR; A57013; A57013;
DR PDB; 1HY1; 14-MAR-01.
DR PDB; 1HYJ; 14-MAR-01.
DR PDB; 1JOC; 28-DEC-01.
DR Genew; HGNC:3185; BEA1.
DR MIM; 605070; -;
DR GO; GO:0005737; C:cytoplasm; TAS.
DR GO; GO:0005769; C:early endosome; TAS.
DR GO; GO:0005624; C:membrane fraction; TAS.
DR GO; GO:0005545; F:phosphatidylinositol binding; TAS.
DR GO; GO:0006899; P:nonspecific vesicle transport; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR000306; Znf_FYVE.
DR Pfam; PF01363; FYVE; 1.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00064; FYVE; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS0178; ZF_FYVE; 1.
DR PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
DR Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
FT DOMAIN 74 1348
FT DOMAIN 397 758
FT DOMAIN 937 1032
FT DOMAIN 1093 1231
FT ZN_FING 41 64
FT ZN_FING 1352 1410
FT MUTAGEN 39 39
FT MUTAGEN 41 41
FT MUTAGEN 42 42
FT MUTAGEN 44 44
FT MUTAGEN 47 47
FT MUTAGEN 60 60
FT MUTAGEN 1349 1349
FT MUTAGEN 1352 1352
FT MUTAGEN 1357 1357
FT MUTAGEN 1358 1358
FT MUTAGEN 1365 1365
FT MUTAGEN 1367 1367
FT MUTAGEN 1370 1370
FT MUTAGEN 1371 1371
FT MUTAGEN 1372 1372
FT MUTAGEN 1373 1373
FT MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN 397 758
DOMAIN 937 1032
DOMAIN 1093 1231
ZN_FING 41 64
ZN_FING 1352 1410
MUTAGEN 39 39
MUTAGEN 41 41
MUTAGEN 42 42
MUTAGEN 44 44
MUTAGEN 47 47
MUTAGEN 60 60
MUTAGEN 1349 1349
MUTAGEN 1352 1352
MUTAGEN 1357 1357
MUTAGEN 1358 1358
MUTAGEN 1365 1365
MUTAGEN 1367 1367
MUTAGEN 1370 1370
MUTAGEN 1371 1371
MUTAGEN 1372 1372
MUTAGEN 1373 1373
MUTAGEN 1374 1374

-!- SIMILARITY: Contains 1 C2H2-type zinc finger.
-!- SIMILARITY: Contains 1 FYVE-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L40157; AAA79121.1; -;
EMBL; X78998; CAA55632.1; -;
PIR; A57013; A57013;
PDB; 1HY1; 14-MAR-01.
PDB; 1HYJ; 14-MAR-01.
PDB; 1JOC; 28-DEC-01.
Genew; HGNC:3185; BEA1.
MIM; 605070; -;
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005769; C:early endosome; TAS.
GO; GO:0005624; C:membrane fraction; TAS.
GO; GO:0005545; F:phosphatidylinositol binding; TAS.
GO; GO:0006899; P:nonspecific vesicle transport; TAS.
InterPro; IPR007087; Znf_C2H2.
InterPro; IPR000306; Znf_FYVE.
Pfam; PF01363; FYVE; 1.
Pfam; PF00096; zf-C2H2; 1.
SMART; SM00064; FYVE; 1.
SMART; SM00355; Znf_C2H2; 1.
PROSITE; PS0178; ZF_FYVE; 1.
PROSITE; PS0028; ZINC_FINGER_C2H2_1; 1.
PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
Antigen; Metal-binding; Zinc; Coiled coil; Zinc-finger; 3D-structure.
DOMAIN 74 1348
DOMAIN

FT MUTAGEN 1375 1375 PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT R->G: ABOLISHES PHOSPHATIDYLINOSITOL 3-
 FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
 FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT G->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
 FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT C->A: ABOLISHES PHOSPHATIDYLINOSITOL 3-
 FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT R->G: STRONGLY REDUCES
 FT PHOSPHATIDYLINOSITOL 3- PHOSPHATE BINDING
 FT AND ABOLISHES ENDOSOMAL LOCATION.
 FT C->S: ABOLISHES PHOSPHATIDYLINOSITOL 3-
 FT PHOSPHATE BINDING AND ENDOSOMAL LOCATION.
 FT C->S (IN REF. 1).
 FT LQ -> FE (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT D -> R (IN REF. 1).
 FT D -> E (IN REF. 1).
 FT EQ -> DE (IN REF. 1).
 FT 575 576

Alignment Scores:

Pred. No.: 5,04e-07 Length: 1411
 Score: 260.50 Matches: 159
 Percent Similarity: 42.46% Conservativity: 134
 Best Local Similarity: 23.04% Mismatches: 232
 Query Match: 7.66% Indels: 165
 DB: 1 Gaps: 33

US-09-602-362E-15 (1-2030) x BEA1_HUMAN (1-1411)

QY 42 ATGAAAGTTCTTATCCAACTAAAGCTTGAATTTG---ATGGACATGCAAACTTTCAA 98
 DB 157 MetLysAspLeuPheGluGlnLysAlaAlaGlnLeuAlaThrGluLeuAlaAspLeuLys 176
 QY 99 GCAGAGCTCCGAGAGCACTCTGCCTTCGAGCCTGCCATTGCAATGCAAAAGTCTGTT 158
 DB 177 SerLysTyrAspGluArgSerLeuArgGluAlaAla-----GluGlnLysVal 193
 QY 159 CCAATAAAGCCTTGGATTGAAGATGAACAAACATTGAGAGCAGATGAGATATCCCA 218
 DB 194 ThrArgLeuThrGluGluLeuAsnLysGluAlaThrValIleGlnAsp-----LeuLys 211
 QY 219 TCAGATCCAAACAAAGAGCTATGAGAAAGTTCTTCGGATTCTCGAGATCTCTGTGAG 278
 DB 212 ThrGluLeuLeuGlnArgProGlyIleGlu----- 221
 QY 279 ACTGTTTACAGAGGATGTGTGTTACCCAGGCTACACATCAAAAGAAATAGATAAA 338
 DB 222 -----AspValAlaValLeuLys-----LysGluLeuValGln 232
 QY 339 ATA-----AATGGAAATTAGAAGAGTCTCTGATAATGATGTTTCTG 383
 DB 233 ValGlnThrLeuMetAspAsnMetThrLeuGluArgGluArgGluSerGluLys---Leu 251
 QY 384 AAGGCTCCCTGC---AGATGAAGTTCTTATCCCACTAAAGCCTTAGAATGATGAC 440
 DB 252 LysAspGluCysLysLysLeuGlnSerGlnTyrAlaSerSerGluAlaThrIleSerGln 271
 QY 441 ATGCAAACTTCAAAGCAGAGCCTCCGAGAGGCCATCTGCTTCGAGCCTGCCATTGAA 500
 DB 272 LeuArgSerGluLeuAlaLysGlyProGlnGluValAlaValTyrValGlnGluLeuGln 291
 QY 501 ATGCAAAAGTCTGTTCCAAATAAAGCCTTGGAAATGAAAGATGAACAAACATTGAGACA 560
 DB 292 LysLeuLysSerSerValAsnGlu---LeuThrGlnLysAsn---GlnThrLeu---Thr 308
 QY 561 GATCAGATGTTCCCTTCAGATCAAAACAAAGAGAGTTGAGAAATCTTCGGATTCT 620
 DB 309 GluAsnLeuLeuLysLysGluGlnAspTyrThrLysLeuGluGlu----- 323
 QY 621 GAGAGTCTCCGTGAGACTGTTTTCAGAGAGGATGTGTGTGATCCCAAGGCTTACA---CAT 677
 DB ---

Db 324 LysHisAsnGluGluSerValSerLysLysAsnIle-----GlnAlaThrLeuHis 340
 QY 678 CAAAGAAGAAATGGAT----- 692
 Db 341 GlnLysAspLeuAspCysGlnGlnLeuGlnSerArgLeuSerAlaSerGluThrSerLeu 360
 QY 693 -----AAATAAGTGGAAATTAAGAATTCAACTAGCTATCAAAATC 737
 Db 361 HisArgIleHisValGluLeuSerGluLysGlyGluAlaThrGlnLysLysGluGlu 380
 QY 738 TTGGATACAGTCTATTCT-----TGTAAGA 764
 Db 381 LeuSerGluValGluThrLysTyrGlnHisLeuLysAlaGluPheLysGlnLeuGln 400
 QY 765 GCAAGGCAACTTCAAAGATCACTGTGAACAACAGTACAGGAAATAATGCAACAATGAAA 824
 Db 401 GlnArgGluGluLysGlnHisGlyLeuGlnLeuGlnSerGluIleAsnGlnLeuHis 420
 QY 825 AAGAAGTTTGTCTACTGAAAAAGAACTCTCAGAGCA-----AAAGAAATA 872
 Db 421 SerLysLeuLeuGluThrGluArgGlnLeuGlyGluAlaHisGlyArgLeuLysGlu--- 439
 QY 873 AAATCAGTTAGAGAACCAAAAA---GTTAAATGGGAACAAGAGCTTGCAGTGTGAGA 939
 Db 440 GlnArgGlnLeuSerSerGluLysLeuMetAspLysGluGlnGlnValAlaAspLeuGln 459
 QY 930 TTGACTTTAAACCAAGAGAGAGAGAGAAATCCGATATATTAATGAAAAAATT 999
 Db 460 LeuLysLeuSerArgLeuGluGlnLeuLysValThrAsnSerThrGluLeu 479
 QY 990 AGGAGAAGAAATTAGAAAGATCGAAGAGCAGCATAGGAAGAGTTAGAAAGTGAACAACAA 1049
 Db 480 GlnHisGlnLeuAspLysThrLysGlnHisGlnGlnGlnGlnGlnGlnSer 499
 QY 1050 CTTGACAGCTCTCAGATACAGATATAGATTGAGAGTGTAGAAAGTAAATTGAAAT 1109
 Db 500 ThrThrAlaLysLeuArgGluAlaGlnAsnAspLeuGlnValLeuArgGlnIleGly 519
 QY 1110 CAGTTTCTCACACTCATGAAATGAAATTTATCTCTTACATGAAATTCATGTTGAAA 1169
 Db 520 AspLysAspGlnLysIleGlnAsnLeuGluAlaLeuLeuLysSer-----Lys 536
 QY 1170 AAGAATAATCCATGCTAAACTGGAATAGCCACACTGAACACCAATACAGGAAAG 1229
 Db 537 GluAsnIleSerLeuLeuGluLysGluArgGluAspLysThrAlaLysIleGlnAlaGly 556
 QY 1230 GAAATAAATACTTTGAGGACATTAAAGATTTTAAAGAAAGAAATGCTGAACCTCAGATG 1289
 Db 557 GluGlyGlu---ThrAlaValLeuAsnGlnGlnGlnLysAsnHisThrLeuGln--- 574
 QY 1290 ACCCTAAACTGAAAGAGGAATCATTAACCTAAAAGGCGATCTCAATATAGTGGCGAGCTT 1349
 Db 574 ----- 574
 QY 1350 AAAGTTCTGATAGTGAACACAAATGCTCAGTTCTTAATTTGAAGGAAACAAAGCAAA 1409
 Db 575 -----GluGlnValThrGlnLeuThrGluLysLeuLysAsnGlnSer----- 588
 QY 1410 GAAATACTAGAGGACAAAATTGAATCACACCATCTAGACTGCTGCTGTACAGAC 1469
 Db 589 -----GluSerHis-----LysGlnAlaGlnGlu 596
 QY 1470 -----CATGATCAAAATTGACATCAAGAAAAGTCAAGAACTCTTCCCATGCA 1523
 Db 597 AsnLeuHisAspGlnVal-----GlnGlnLysAlaHisLeuArg 610
 QY 1524 CGAGATGCTTTGTCAAAAGAAAATGAATGTTGATGTAGTAGTACCATATATACAA 1583
 Db 611 -----AlaAlaGlnAspArgValLeuSerLeuGluThrSerValAsnGluLeuAsnSe 628
 QY 1584 TGAGGTGCTCCATCAACCACTTCTGAAGCTCAAGGAAATCCANAAGCCTTAAATTTAA 1643
 Db 628 rGln-----LeuAsnGluSerLysGluLysValSerGlnLeuAspIleG 643

Db 839 ThrLysValLysProLeuLeuGlnValThrArgGlnGluGluMetGlnAlaLysGlu 858
QY 477 TCTGCTTCGAGCTGCCATTGAATGCAAAAGTCTGTTCCTCAAAATAAGACCTTCGAATTG 536
Db 859 GluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThrGluLeuLysGluLeu 878
QY 537 AAGAAATGAACAAACATTCAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAG 596
Db 879 GluGlnLysHisThrGlnLeuAlaGluLysThrLeuLeuGlnGlnGlnAla 898
QY 597 GTTGAAGAAATCTTCGGATTCTCAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTG 656
Db 899 GluThrGluLeuTyraGluSerGluMetArg 910
QY 657 TGTGTACCAAGCTACATCAAAAGAAATGGATAAAATTAAGTGGAAATTAAGAGAT 716
Db 911 ---ValArgLeuAlaAlaLysLysGlnGluLeuGluGluIle--- 923
QY 717 TCACTAGCTATCAAAATCTTGATACAGTTCATCTTCTGAAAGAGCAGGGAAGCTT 776
Db 924 ---LeuHisGluMetGluAlaArgLeuGluGlu 933
QY 777 CAATAAGATCACTGTGAACAACGTACAGGAAAAATGGAACAAATGGAAGAAAGCTTTGT 836
Db 934 GluGluAspArgArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeu 953
QY 837 GTACTGAAAGAAATCTGTCAAGCAACAAAGAAATPAAATCACAGTTA 884
Db 954 AspLeuGluGlnLeuGluGluGluAlaAlaArgGlnLysLeuGlnLeuLys 973
QY 885 ---CAG 887
Db 974 ValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAspGln 993
QY 888 AACCAAAAGTTAAATGGGAACAAGAGCTGTGACGTGTGAGATTGACT 935
Db 994 AsnSerLysLysSerLysGluArgLysLeuLeuGluGluArgValSerAspLeuThr 1013
QY 936 ---TTAAACCAAGAGAGAGAGAGAGAAATGCCGATATATTAATAAGAAAAATAGG 992
Db 1014 AsnLeuAlaGluGlnGluLysAlaLysAsnLeuThrLysLeuLysSerLysHisGlu 1033
QY 993 GAAGAATTAGGAAGAAATCAAGAGCAGCATAGGAAGAGTTAGAGTGAACCAACAACT 1052
Db 1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluGluLysSerArgGlnGluLeu 1053
QY 1053 GAACAGCTCTCAGATACAGATATAGATTGAGAGCTGAGAAATTAATTGAAATCAG 1112
Db 1054 GluLysLeuLysArgLysLeuGlu 1063
QY 1113 GTTCTCAGCTCATGAAATGAAATATTCTCTTACATGAAATTCATGTTGAAAAAG 1172
Db 1064 AlaSerAspPheHisGlu 1069
QY 1173 GAAATGGCATGCTAAACTGAATAGCCACACTGAAACACCAATACAGAGAAAGAA 1232
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
QY 1233 AATAAATCTTTGAGCAGATTAAATTTTAAAGAAAGAAATGCTGAACTTCAGATGACC 1292
Db 1090 GluGluLeuGlnAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAsnAla 1109
QY 1293 CTAATAA 1319
Db 1110 LeuLysLysIleArgGluLeuGluGlyHisIleSerAspLeuGlnGluAspLeuAsp 1129
QY 1320 AAAAGGCTATCAATATAGTGGG 1352
Db 1130 GluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGlu 1149
QY 1353 GTTCTGATAGCTGAG 1394
Db 1150 AlaLeuLysThrGluLeuGluAspThrLeuAspSerThrAlaThrGlnGlnLeuArg 1169

RESULT 15

USOL_YEAST STANDARD; PRT; 1790 AA.
ID USOL_YEAST
AC P25386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Intracellular protein transport protein USOL.
GN USOL OR INTL OR YDLO58W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol1, is required for intracellular

QY 1395 GAAAAACAAGACAAAGAAATA 1436
Db 1170 AlaLysArgGluGlnGluValThrValLeuLysLysAlaLeuAspGluThrArgSer 1189
QY 1437 CACCATCTAGAGTCTTCTGCTGTACAGACCATGATCAAAATTCGATCATCAAGAAAA 1496
Db 1190 HisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThr 1209
QY 1497 AGTCAAGAACCTCTTTCCACATTCGAGAGATGCTTGTTCGAAAGAAAAATGAATGTT 1556
Db 1210 GluGlnLeuGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeu 1229
QY 1557 GATGTGAGTAGTACCATATATAACAATGAG 1610
Db 1230 GluLysGluAsnAlaAspLeu-AlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnGlu 1249
QY 1611 AGCTCAAGGAAATCCANAAGCCTAAAAATTTATCTC 1658
Db 1249 uValGluHisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysCysSe 1269
QY 1659 AGATGCTTAAAGAAAAATACATTTGTTTCAGGACATGCACAAAGACCAACCTGAAA 1718
Db 1269 rAspGlyGluArgAlaArgAlaGluLeuSer-AspLysValHisLysLeuGlnAsnGluV 1289
QY 1719 CACAGTGT 1744
Db 1289 alGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysA 1309
QY 1745 ---CACATGNTCAAACGAAACAGATNATGTGAAACAAACACACTG 1787
Db 1309 spValAlaSerLeuGlySerGlnLeuGlnAspThrGlnLeuLeuGlnGluGluThrA 1329
QY 1788 ANCAGCAGGAGTCTTAGATCAGAAATTTTCAACTCAACAAACAAACAAATATGCGCTTC 1847
Db 1329 rgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuG 1349
QY 1848 AACAGCAATTAGTTCATGCACATAANGAAGCTGCACACAAACAAACAAACAAATGAT 1907
Db 1349 iAspGlnLeu 1363
QY 1908 NTTCATNTTCTGAGAGGAAATGCNCATCATCTTTAAAGAGAAAAATGAGAGATAT 1967
Db 1364 ArgHisValSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPhe 1383
QY 1968 TTNATTACNATAACCATTTAAAAAACCCGTATATTTCAATATATGAAAAAANAANA 2027
Db 1384 AlaSerThrIle 1398
QY 2028 AAA 2030
Db 1399 Lys 1399

RT protein transport in Saccharomyces cerevisiae. J. Cell Biol. 113:245-260(1991).
 RL [2]
 RN SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RL Kendrick K.E.;
 RA Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Required for protein transport from the ER to the Golgi
 CC complex.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, composed
 CC of an heptapeptide repeat pattern characteristic of alpha-helical
 CC coiled coils. May form filamentous structures in the cell.
 CC -1- SIMILARITY: BELONGS TO THE VDP/USO1/YBL047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X54378; CAA38253.1; -;
 DR EMBL; L03188; AAB00143.1; -;
 DR EMBL; U53688; AAB66659.1; -;
 DR GenBank; U53688; AAB66659.1; -;
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR02017; Spectrin.
 DR InterPro; IPR006955; USO1 p115 C.
 DR InterPro; IPR006953; USO1 p115_head.
 DR Pfam; PF04871; USO1 p115_C; 1.
 DR Pfam; PF04869; USO1 p115_head; 1.
 DR Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 FT DOMAIN 1 724 GLOBULAR HEAD.
 FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
 FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 FT CONFLICT 847 847 G -> E (IN REF. 2).
 FT CONFLICT 924 924 E -> K (IN REF. 2).
 FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 FT CONFLICT 1772 1772 D -> DEEDDE (IN REF. 2).
 SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Alignment Scores:
 Pred. No.: 7,92e-07 Length: 1790
 Score: 256.50 Matches: 163
 Percent Similarity: 39.45% Conservative: 136
 Best Local Similarity: 21.50% Mismatches: 298
 Query Match: 7.54% Indels: 161
 DB: 1 Gaps: 28

US-09-602-362E-15 (1-2030) x USO1_YEAST (1-1790)

QY 51 TCATTCCTCAACT-----AAAGCCTTAGAATGATGAGCATGCAACACTTTCACACGAG 104
 DB 840 SerileLeuThrLeuGluLysGlyLeuGluThrIleLeuSerGlnLysLysLysAlaGlu 859
 QY 105 CCT-----CCUGAGAAGCCATCTGCTTCGAGCCTGCCATTGAAATGCAAAAGTCT 155

Db AspGlyIleAsnLysMetGlyLysAspLeuPheAlaLeuSerArgGluMetGlnAlaVal 879
 QY 156 GTTCCAAATAAGCCTTGGAAATGAAGAATGAACAAACATTGAGAGCAGATGAGATAC 215
 Db GluGluAsn----- 882
 QY 216 CCATCAGATCCAAACAAAGAGACTATGAAGAAGTTCTTGGGATCTGAGACTCTCTGT 275
 Db 883 ---CysLysAsnLeuGlnLysGluLysAspLysSerAsnValAsnHisGlnLys----- 899
 QY 276 GAGACTCTTTCCAGAGAGGATGTGTTTACCAAGGCTACACATCAAAAGAAATAGAT 335
 Db 900 GluThrLysSerLeuLysGluAspIleAlaLysIleThr-----GluIleLys 916
 QY 336 AAAATAATGAAAAATTAGAAAGATCTCTCTGATAATGATGTTTCTGAAGGCTCCCTGC 395
 Db 917 AlaIleAsnGluAsnLeuGlu-----MetLysIleGlnCys 929
 QY 396 -----AGATGAAAGTTTCTATTCACACTAAAGCCTTAGAATGTGAGCATG 443
 Db 930 AsnAsnLeuSerLysGluLysGluHisIleSerLysGluLeuValGlu----- 945
 QY 444 CAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCTTCGAGCCTGCCATTGAAATG 503
 Db 946 -----TyrLysSerArgPheGlnSerHisAspAsnLeuValAlaLysLeuThrGluLys 963
 QY 504 CAAAGCTCTTCCAAAT-----AAAGCCTTGAATGAAGATGAACAAACATTGAGA 557
 Db 964 LeuLysSerLeuAlaAsnAsnTyrLysAspMetGlnAlaGluAsnGluSerLeuLys 983
 QY 558 GCAGATCAGATGTTCCCTTCAGAAATCAAAA-----CAAAAGAGGTT 599
 Db 984 AlaValGluLysLysAsnGluSerSerIleGlnLeuSerAsnLeuGlnAsnLysIle 1003
 QY 600 GAAGAAATTTCTGGGATCTTGAGAT----- 626
 Db 1004 AspSerMetSerGlnGluLysGluAsnPheGlnIleGluArgGlySerIleGluLysAsn 1023
 QY 627 -----CTCCGTGAGACTGTTTTCACAG-----AAGGATGCTGTGTACCC 665
 Db 1024 IleGluGlnLeuLysLysThrIleSerAspLeuGluGlnThrLysGluLysIleSer 1043
 QY 666 AAGGCT-----ACACATCAAAAGAAATGATAAATAAGTGAAATTA 710
 Db 1044 LysSerAspSerSerLysAspGluTyrGluSerGlnIleSerLeuLysGluLysLeu 1063
 QY 711 GAAGATCAACTAGC-----CTATCAAAATCTTGGATACA 746
 Db 1064 GluThrAlaThrThrAlaAsnAspGluAsnValAsnLysIleSerGluLeuThrLysThr 1083
 QY 747 GTTCATTCTTGTGAAAGAGCAGAGGAACTCAAAAAGATCACTGTGCAACACGTACAGGA 806
 Db 1084 ArgGluGluLeuGluAlaGluLeuAlaAlaTyrLysAsnLeuLysAsnGluLeuGluThr 1103
 QY 807 AAAATGCAACAAATGAAAGAGGTTTGTGACTGAAAGAAAGAACTGTGCAAGCAAAA 866
 Db 1104 LysLeuGluThrSerGluLysAlaLeuLysGluValLysGluAsnGluGluHisLeuLys 1123
 QY 867 GAAATAAATTCAGATTAGAACCAACAAAGTTTAAATGGGAACAGAGCTCTGCACTGTG 926
 Db 1124 GluGluLysIleGlnLeuGluLysGluAlaThrGluThrLysGlnGlnLeuAsnSerLeu 1143
 QY 927 AGATTGACTTTAAACCAAG 986
 Db 1144 ArgAlaAsnLeuGluSerLeuGluLysGluHisGluAspLeuAlaAlaGlnLeuLysLys 1163
 QY 987 ATTAGGGAAGATTAGGAGATCGAAGAGCAGCATGAGGAGAGAGAGAGAGAGAGAGAGAG 1037
 Db 1164 TyrGluGluGlnIleAlaAsnLysGluArgGlnTyrAsnGluLysIleSerGlnLeuAsn 1183
 QY 1038 -----GTGAAACCAACAACTTGAACAGGCTCTCAGAAATACAGATATAGATTA 1085
 Db 1184 AspGluIleThrSerThrGlnGlnGluAsnGluSerIleLysLysLysAsnAspGluLeu 1203

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 09:09:52 ; Search time 29.6705 Seconds
(without alignments)
13162.486 Million cell updates/sec

Title: US-09-602-362E-15
Perfect score: 3403
Sequence: 1 ctctgctgcttaagatgggt.....gaaaaaanaaaaaaaa 2030

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 56732
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p model -DEV=xlp
-O=/sqn2.1/USPTC.spool_P/US09602362/runat.15072004.093626.22015/app_query.fasta.1.10325
-DB=PIR78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEL=0 -LDOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09602362 @CN 1.1.248 @runat.15072004.093626.22015 -NCPU=6 -ICPU=3
-NO MWAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR 78:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	318.5	9.4	1738	T14867	interaptin - slime
2	280.5	8.2	1875	S38173	myosin-like protei
3	280	8.2	2057	S61477	myosin II heavy ch
4	280	8.2	2116	A26655	myosin heavy chain
5	279	8.2	2017	A36014	myosin heavy chain
6	278.5	8.2	1642	T08880	NMDA receptor-bind
7	275.5	8.1	1690	T13030	microtubule bindin
8	267	7.8	1979	C71622	hypothetical prote
9	267	7.8	2954	T14156	kinesin-related pr
10	266.5	7.8	2663	S28261	centromere protein
11	266	7.8	944	S26710	spindle pole body
12	262	7.7	1388	S70633	serine/threonine-s
13	261.5	7.7	1972	A41604	myosin heavy chain
14	261	7.7	2007	B43402	myosin heavy chain

15	258	7.6	1938	2	JC5421	smooth muscle myos
16	258	7.6	1972	2	JC5420	smooth muscle myos
17	256.5	7.5	1790	2	S67593	transport protein
18	255.5	7.5	2442	2	T08621	cytosome associa
19	255	7.5	1979	1	S03166	myosin heavy chain
20	255	7.5	2101	2	A42184	nuclear mitotic ap
21	254	7.5	1961	1	A61231	myosin heavy chain
22	254	7.5	2245	2	T18278	myosin heavy chain
23	253	7.4	1939	2	T48175	myosin heavy chain
24	253	7.4	2253	2	T30336	nuclear/mitotic ap
25	252	7.4	1939	1	A46762	myosin alpha prote
26	250.5	7.4	992	2	T46337	hypothetical prote
27	250.5	7.4	1388	2	S74245	serine/threonine-s
28	249	7.3	1938	1	S06005	myosin alpha heavy
29	248.5	7.3	853	2	T51505	hypothetical prote
30	248.5	7.3	2168	2	T30171	ninin - mouse
31	247.5	7.3	3488	2	T34418	hypothetical prote
32	247	7.3	1938	2	I49464	alpha cardiac myos
33	247	7.3	1976	2	A59252	myosin heavy chain
34	247	7.3	3259	1	A56539	giantin - human
35	246	7.2	924	2	S06117	myosin heavy chain
36	246	7.2	3225	2	I52300	giantin - human
37	245	7.2	1410	1	A57013	early endosome ant
38	244	7.2	1130	2	T34081	hypothetical prote
39	244	7.2	1295	2	T24587	hypothetical prote
40	243.5	7.2	1935	1	A37102	myosin beta heavy
41	243	7.1	1085	2	F96712	hypothetical prote
42	243	7.1	1999	1	S21801	myosin heavy chain
43	242.5	7.1	1392	2	A43336	microtubule-vesicl
44	242	7.1	1269	2	F84730	probable myosin he
45	242	7.1	3187	2	JC5837	364K Golgi complex

ALIGNMENTS

RESULT 1

T14867
interaptin - slime mold (Dictyostelium discoideum)
C/Species: Dictyostelium discoideum
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C/Accession: T14867
R/Roverto, P.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A/Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dictyo
ts.
A/Reference number: Z18248; MUID:98365468; PMID:9700162
A/Accession: T14867
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1738 <RIV>
A/Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1
C/Genetics:
A/Gene: abpD
A/Introns: 173/2; 1680/1

Alignment Scores:
Pred. No.: 6,89e-12 Length: 1738
Score: 318.50 Matches: 180
Percent Similarity: 37.86% Conservative: 146
Best Local Similarity: 20.91% Mismatches: 244
Query Match: 9.36% Indels: 291
DB: 2 Gaps: 33

US-09-602-362E-15 (1-2030) x T14867 (1-1738)

QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCGAGAGCATCTGCCTTC 128
|||||
585 LeuGlnLeuGlnAspGlnAspIleLys-----GluLysGluPheGlnPhe 600
|||||
QY 129 GAGCCT-----GCCATTGAATGCAA 149
|||||
601 GluLysGlnGlnLeuLeuSerGlnIleAspSerIleThrThrAsnIleGlnLutryGln 620

QY 357 ---GAGTCTCCTGAT-----AAT 371
 Db 1315 AlaGluAsnAlaAspLeuAlaThrGluLeuArgSerValAsnSerSerArgGlnGluAsn 1334
 QY 372 GATGGTTTCTGAAGGCTCCCTGCAGAAATGAAGTTCTTATTCCAACTAAAGCCCTTAGAA 431
 Db 1335 AspArgArgGlys-----GlnAlaGluSerGlnIleAlaGluLeuGlnVallys 1351
 QY 432 TTGATGGACATCAAACTTTCAAGCAGAGCCCTCCGAGAACCCATCTGCTTCGAG--- 488
 Db 1352 LeuAlaGluIleGluArgAlaArgSerGluLeuGlnGluLysCysThrLysLeuGlnGln 1371
 QY 489 -----CCTGCGATTGAATGCCAAATGCCAAAGCTCTGTTCCAAAT 521
 Db 1372 GluAlaGluAsnIleThrAsnGlnLeuGluGluAlaGluLeuLysAlaSerAlaAlaVal 1391
 QY 522 AAAGCCTTGGAAATGAAGAAATGAACAAACATTGAGAGCAGATCAGATGTTCCCTTCAGAA 581
 Db 1392 LysSerAlaSerAsnMetGluSerGlnLeuThrGluAlaGlnGlnLeuLeuGluGluGlu 1411
 QY 582 TCANAACAAAG-----RAGGTTGAAGAAATCTTGGATCTCGAGAT 626
 Db 1412 ThrArgGlnLysLeuGlyLeuSerSerLysLeuArgGlnIleGluSerGluLysGluAla 1431
 QY 627 CTCCGTGAGACTGTTTCACAGAGAGGATGTGTGTACCCAGGCTACACATCAAAAA--- 683
 Db 1432 LeuGlnGluGlnLeuGluAspAspGluAla-----LysArgAsnTyrGluArgLys 1449
 QY 684 -----GAAATGATAAATACTGGAATAATTAAGAAATTAAGAGATTTCAACTAGC 725
 Db 1450 LeuAlaGluValThrGlnMetGlnGlnIleLysLysLysAlaGluGluAspAlaAsp 1469
 QY 726 CTATCAAAAATCTTGATACAGTTTCATCTTGTGAAGAGCAGGAGGAACTTCAAAAGAT 785
 Db 1470 LeuAlaLysGluLeuGlu-----GluGlyLysLysArgLeuAsnLysAsp 1484
 QY 786 CACTGTGAACACGTACAGGAAAATGAACAAATGAATAAGTATTGTTGTACTGAAA 845
 Db 1485 -----IleGluAlaLeuGluArgGlnValLysGluLeuIleAlaGlnAsn 1499
 QY 846 AAGAACTGTCAGAGCA---AAGAAATAAATCACAGTTAGAGAACCAACAAAGTTAAA 902
 Db 1500 AspArgLeuAspLysLysLysLysLysIleGlnSerGluLeuGluAspAlaThrIleGlu 1519
 QY 903 TGGGAACAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGA 962
 Db 1520 LeuGluAlaGln-----ArgThrLysValLeuGluLeuGluLysLysGlnLys 1535
 QY 963 AATGCCGAT-----ATATTAAATGAATAAATTAGGAGAAATTA 1001
 Db 1536 AsnPheAspLysIleLeuAlaGluGluLysAlaIleSerGluGlnIleAlaGlnGluArg 1555
 QY 1002 GGAAGAAATCGAA-----CAGCAGCATAGGAAAGAGTTAGAGTCAAAACAA 1049
 Db 1556 AspThrAlaGluArgGluAlaArgGluLysGluThrLysValLeuSerValSerArgGlu 1575
 QY 1050 CTTGAACAGGCTCTC---AGAATCAAGATATAGAA-----TTGAGAGGTGTAGAAAGT 1100
 Db 1576 LeuAspGluAlaPheAspLysIleGluAspLeuGluAsnLysArgLysThrLeuGlnAsn 1595
 QY 1101 AATTTTGAATCAGGTTTCTCACACTCATGAAATGAATAATTTATCTTACATGAAATTCG 1160
 Db 1596 GluLeuAspAspLeuAlaAsnThrGlnGlyThrAlaAspLysAsnValIleGlu----- 1613
 QY 1161 ATGTTGAAAGGAAATTCGCCATGCTAAATCTGGAATAGCCACACTGAAACACCAATAC 1220
 Db 1614 ---LeuGluLysAlaLysArgAlaLeuGluSerGlnLeuAlaGluLeuLysAlaGlnAsn 1632
 QY 1221 CAGGAAAGGAAAT-----AAATACCTTTGAGGACATTAGATTTTAAACAAAGAAAT 1274
 Db 1633 GluGluLeuGluAspAspLeuGlnLeuThrGluAspAlaLysLeuArgLeuGluValAsn 1652

QY 1275 -----GCTGAACCTTCAGATGACCTTAAACTGAAGAGGAAATCATTAAC 1319
 Db 1653 MetGlnAlaLeuArgSerGlnPheGluArgAspLeuLeuAlaLysGluGluGlyAlaGlu 1672
 QY 1320 AAAAGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGTAGTCTGAG----- 1367
 Db 1673 GluLysArgArgGlyLeuValLysGlnLeuArgAspLeuGluThrGluLeuAspGluGlu 1692
 QY 1368 -----AACCAATGCTCCTTAAATTGAAGAAAACAAAGAAAGAAATA 1415
 Db 1693 ArgLysGlnArgThrAlaAlaValAlaSerLysLysLysLeuGluGlyAspLeuLysGlu 1712
 QY 1416 CTAGAGCAGCAAAATGAAATCACACCATCTAGACTGCTTCTGCTGTACAGACCATGAT 1475
 Db 1713 IleGluThrThrMetGluMetHisAsnLysValLysGluAspAlaLeu-LysHisAlaLys 1732
 QY 1476 CAAATTTGACATCAAGAAAAGTCAAGAACCTGCTTCCACATTCAGAGAGATGCTTGT 1535
 Db 1732 sLysLeu-----GlnAlaGlnValLysAspAlaLeuArgAspAlaGluGluAlaLysAl 1750
 QY 1536 TTGCAAGAAAATGAATGTTGATG-----TGAGTAGTACCGATATATA 1580
 Db 1750 aAlaLysGluGluLeuGlnAlaLeuSerLysGluAlaAspGlyLysValLysAlaLeuGlu 1770
 QY 1581 CAATGAGTCTCTCCAT-----CAACACCTTCTGAGGCTCAAGGAAATCCANAAGCCT 1634
 Db 1770 uAlaGluValLeuGlnLeuThrGluAspLeuAlaSerSerGluArgAlaArgAlaAl 1790
 QY 1535 AAAAATTATCTCAATATGCGAGAGATGCTCTAAGAGAAAATACATTTGTTTCAGAAC 1694
 Db 1790 aGluThrGluArgAspGluLeuAlaGluIleAlaAsnAsnAlaAsnLysGlySerLe 1810
 QY 1695 ATGCACAAAGAGACCAACGTAACACAGTGTCAATGAAGAAAGCTGACACATGNTC 1754
 Db 1810 uMetIleAsp-GluLysArgArgLeuGluAlaArgIleAlaThrLeuGluGluGluLeuG 1830
 QY 1755 AAANCGAACAA-----GATNATGTGAACAAACACACACTGANCAGC 1793
 Db 1830 luGluGluGlnSerAsnSerGluValLeuLeuAspArgSerArgLysAlaGlnLeuGlnI 1850
 QY 1794 AGGAGTCTGATCAGATCAGAAATTTCACTACAAACCAAAAATATGTGCTTCAACAGC 1853
 Db 1850 leGluGlnLeuThrGluLeu----- 1857
 QY 1854 AATTAGTTCATGCACATAAAGAAAGCTGACAAACAAAGCAAGATAACAATTTGATNTTCAT 1913
 Db 1858 -----AlaAsnGluLysSerAsnSerGlnLys----- 1866
 QY 1914 TTTCTTGAGGAGAAATGCNCATCATCTTCTTAAAGAGAAAATAGAGAGATA 1966
 Db 1867 -----AsnGluAsnGlyArgAlaLeuLeuGluArgGlnAsnLysGluLeu 1881
 RESULT 4
 A26655
 myosin heavy chain [similarity] - slime mold (Dictyostelium discoideum)
 N/Contains: myosin ATPase (EC 3.6.4.1)
 C/Species: Dictyostelium discoideum
 C/Date: 05-Oct-1988 #sequence revision 05-Oct-1988 #text_change 19-Apr-2002
 C/Accession: A26655; A24728; S00250
 R/Warrick, H.M.; De Lozanne, A.; Levinwand, L.A.; Spudich, J.A.
 Proc. Natl. Acad. Sci. U.S.A. 83, 9433-9437, 1986
 A/Title: Conserved protein domains in a myosin heavy chain gene from Dictyostelium disco
 A/Reference number: A26655; MUID:87092266; PMID:3540939
 A/Accession: A26655
 A/Molecule type: DNA
 A/Molecule type: DNA
 A/Residues: 1-2116 <WAR>
 A/Cross-references: GB:M14628; GB:M11938; NID:G167834; PID:AAA33227.1; PID:G167835
 R/DeLozanne, A.; Lewis, M.; Spudich, J.A.; Levinwand, L.A.
 Proc. Natl. Acad. Sci. U.S.A. 82, 6807-6810, 1985
 A/Reference number: A24728; MUID:86016788; PMID:3901008
 A/Accession: A24728
 A/Molecule type: mRNA
 A/Residues: 2035-2116

R:Wagle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
 PDB: 1ett, 227, 71-75, 1988
 A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostelium
 A:Reference number: S00250; MUID:188112226; PMID:2828113
 A:Accession: S00250
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1734-1993 <WAG>
 C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: actin binding; ATP; coiled coil; hydrolase; nucleotide binding; P-loop; phosphatase
 F:1-818/Domain: globular head <HD>
 F:89-747/Domain: myosin motor domain homology <MMOT>
 F:179-186/Region: nucleotide-binding motif A (P-loop)
 F:819-2116/Domain: alpha-helical rod <ROD>

Alignment Scores:
 Pred. No.: 1.43e-09 Length: 2116
 Score: 280.00 Matches: 158
 Percent Similarity: 36.61% Conservative: 110
 Best Local Similarity: 21.58% Mismatches: 256
 Query Match: 8.23% Indels: 208
 DB: 2 Gaps: 27

US-09-602-362E-15 (1-2030) x A26655 (1-2116)

QY	24	CTGAAGCGTAACTGCGGAATGAAGTTCTATTCCAACTTAAAGCCTTGAATG---ATG 80	
Db	879	LeuLysAlaMetTyrAspSerLysAspAlaLeuGluAlaGlnLysArgGluLeuGluile 898	
QY	81	GACATCAAACTTCAAAGCAGAGCTCCCGAGAGCCATCTGCCTTCGAGCTCCCAT 140	
Db	899	ArgValGluAspMetGluSerGluLeuAspGluLysLeuAlaLeuGlnLeuGln 918	
QY	141	GAATGCAAAAGTCTGTCCTCAATAAAGCCTTG-----CAATTGAAGAAATGAA 188	
Db	919	AsnGlnLysArgSerValGluLysValArgAspLeuGluGluGluLeuGluGlu 938	
QY	189	CAACATTTGAGCAGATGATGATCTCCCATCAGATCCAAACAAAGAGCTATCAAGAA 248	
Db	939	GlnLysLeuArgAsn-----ThrLeuGluLysLeuLysLysLysTyrGluGlu 954	
QY	249	AGTCTCTGGAT-----TCTGAGAGTCTCTGTCGACACTGTTTCACAGAGGAT 296	
Db	955	GluLeuGluGluMetLysArgValAsnAspGlyGlnSerAspThrIleSerArgLeuGlu 974	
QY	297	GTGTGTTTACCAGCTACATCAAAAGAAATGATATAAATGCGAAATTA--- 353	
Db	975	Lysile-----LysAspGluLeuGlnLysGluValGluGluLeuThrGluSerPheSer 992	
QY	354	GAAGAGTCTCTGATATGATGTTTCTGGAAGCTCCCTGCGAGATGAAGTTTCTATT 413	
Db	993	GluGluSerLysLysGlyValLeuLysThrArgValArgLeuGlnSerGluLeu 1012	
QY	414	CCAACATAAGCCCTTGAATTTGATGACATGCAAACTTTCAAACGACAG----- 461	
Db	1013	AspAspLeuThrValArgLeuAspSerGluThrLysAspLysSerGluLeuLeuArgGln 1032	
QY	462	-----CCTCCGAGAGCCATCTGCTCCCTTCGAGCTGCC 494	
Db	1033	LysLysLysLeuGluGluLeuLysGlnValGlnGluAlaLeuAlaGluThrAla 1052	
QY	495	ATTGAAATGCAAAAGTCTGTTCCAAATAAAGCCTTG----- 530	
Db	1053	AlaLysLeuAlaGlnGluAlaAlaAsnLysLysLeuGlnGlyGluTyrThrGluLeuAsn 1072	
QY	530	----- 530	
Db	1073	GluLysPheAsnSerGluValThrAlaArgSerAsnValGluLysSerLysThrLeu 1092	
QY	531	-----GAATTGAAGAAATGAACAAACA 551	
Db	1093	GluSerGlnLeuValAlaValAsnAsnGluLeuAspGluGluLysLysAsnArgAspAla 1112	

QY	552	TTGAGA-----GCAGATCAGATGTTCT 572	
Db	1113	LeuGluLysLysLysLysAlaLeuAspAlaMetLeuGluGluMetLysAspGlnLeuGlu 1132	
QY	573	CCTTCAGATCAAAACAAAG-----ARGTTCAAGAAATCTCTGGAT 617	
Db	1133	SerThrGlyGluLysLysSerLeuTyrAspLeuLysValLysGlnGluSer---ASP 1151	
QY	618	TCTGAGAGTCTCCGTCAGAGCTGTTTCACAGAGGATGTGTGTACCC----- 665	
Db	1152	MetGluAlaLeuArgAsnGlnIleSerGluLeuGlnSerThrIleAlaLysLeuGluLys 1171	
QY	666	---AAGCTACATCAAAAGAAATCGATATAAATAGTGGAAATAGAAATCAACT 722	
Db	1172	IleLysSerThrLeuGluGluValAlaLeuArgLeuGlnGlyGluLeuGlu----- 1188	
QY	723	AGCCTATCAAAATCTTGGATACAGTTTCATCTTGTGAAAGAGCAAGGAACTTCAAAA 782	
Db	1189	-----AlaGluGlnLeuAlaLys 1194	
QY	783	GATCAGTGTGAACACAGTACAGGAAATCGAA---CAAATGAAAGAAAGTTTCTGTA 839	
Db	1195	SerAsnValGluLysGlnLysLysValGluLeuAspLeuGluAspLys----- 1211	
QY	840	CTGAAAGAAAGCTGTCAAGCAAGAAATAAATCACAGTTAGAGAACCAAAAAGTT 899	
Db	1212	---SerAlaGlnLeuAlaGluThrAlaAlaLysGlnAlaLeuAspLysLysLys 1230	
QY	900	AAATGGGAACAAGCTCTGACGTGTGACATTTAAACCAAGCAAGAGAGAGA 959	
Db	1231	LysLeuGluGlnLeuSerGluValGlnThrGlnLeuSerGluAlaAsnAsnLysAsn 1250	
QY	960	AGAAATCCCATATATTAATGAAATAATAGGAAGATTAGGAAGATTCGAAGAGCAG 1019	
Db	1251	ValAsnSerAspSerThrAsnLysHisLeuGluThrSerPheAsnAsnLeu----- 1267	
QY	1020	CATAGCAAGAGTTAGAAAGTGAACCAACACTTGAACAGCTCTCAGATACAGATATA 1079	
Db	1268	---LysLeuGluLeuGluAlaGluGlnLysAlaLysGlnAlaLeuGluLysLysArgLeu 1286	
QY	1080	GAATTGAAGAGTGAAGAAATTAATTCAGATTCACAGTTCTCAGACTCATGAAATGAAAT 1139	
Db	1287	GlyLeu-----GluSerGluLeuLysHisValAsnGluGlnLeuGluGluLys 1303	
QY	1140	TATCTCTTACATGAAATTCATGTTGAAAGAGAAATTCGCCATGCTAAACTGGAATA 1199	
Db	1304	LysGlnLysGluSerAsn-----GluLysArgLysValAspLeuGluLysGluVal 1320	
QY	1200	GCCACACTGAACACCAATACAGGAA----- 1226	
Db	1321	SerGluLeuLysAspGlnIleGluGluValAlaSerLysLysAlaValThrGluAla 1340	
QY	1227	-----AAGGAAAT-----AAATACTTTCAGGACATTAAGATT 1259	
Db	1341	LysAsnLysLysGluSerGluLeuAspGluLysArgGlnTyrAlaAspValValSer 1360	
QY	1260	TTAAAGAAAGAAATGCTGAACCTTCAGATGACCTAAACTGAAAGAGGATCATTAAT 1319	
Db	1361	SerArgAspLysSerValGluGlnLeuLysThrLeuGlnAlaLysAsnGluLeuArg 1380	
QY	1320	AAAAGGCATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACATGCTC 1379	
Db	1381	AsnThrAlaGluGluAlaGluGlyGlnLeu----- 1390	
QY	1380	ACTTCTAAATTGAAGGAAACCAAGCAAAATACTAGAGGACAGAAATGAATGCAC 1439	
Db	1391	-----AspArgAlaGluArgSerLysLysLysAlaGluPheAsp----- 1403	
QY	1440	CATCTAGACTGGCTTCTGCTGTACAGACCATCATCAATTTGACATCAAGAAAAAGT 1499	
Db	1404	-----LeuGluGluAlaValLysAsnLeuGluGluThrAlaLysLysValLys 1420	

```

QY 1500 CAAGAACCTGCTTTCCACATTGCAGGAGATGCTTGTTCGAAGAAAAATGATGTTGAT 1559
Db 1421 AlaGluLysAla-----MetLysLysAlaGluThrAsp 1431
QY 1560 GTGAGTAGTACCGATATATAACAATGAGGTGCTCCATCAACACTTCTCAAGCTCAAG 1619
Db 1432 TyrArgSerThr-----LysSer 1437
QY 1620 GAATCCANAAGCTCAAAATTAATCTCAATTATGAGGAGATGCTCTAAGAGAAAATAC 1679
Db 1438 GluLeuAspAlaLysAsn-----ValSerSerGluGlnTyr 1450
QY 1680 ATTGGTTTCAGGAACATCACAAAGAGACCAACGTCGAACACAGTGTCAAAATCAAGGAAG 1739
Db 1451 ValGlnIleLysArgLeuAsnGluLeuSerGluLeuArgSerValLeuGluGluAla 1470
QY 1740 CTGAACACATGTTTCAAAACGACACAGATNATGTGAACAAACACACTGANCAGCGAGGT 1799
Db 1471 AspGluArgCysAsnSerAlaIleLysAlaLysLysThrAlaGluSerAlaLeuGluSer 1490
QY 1800 CTCTAGATCAGAAATATTTCAACTACAAGCAAAATATGTGCTTCAACAGCAATTAG 1859
Db 1491 Leu-----LysAspGluIleAspAlaAlaAsnAsn----- 1500
QY 1860 TTCATGCACATANGAAAGCTGACACAAACAAAG 1895
Db 1501 -----AlaLysAlaLysAlaGluArgLysSerLys 1510

RESULT 5
A36014
Mycosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Drosophila melanogaster
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A36014; B36014
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript: cDNA
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <KET>
A:Cross-references: GB:M35012
C:Genetics:
A:Gene: FlyBase:zfp
A:Cross-references: FlyBase:FBgn0005634
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:1-2017/Product: myosin heavy chain, form I #status predicted <MA1>
F:46-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F:135-815/Domain: myosin motor domain homology <MMOT>
F:223-232/Region: nucleotide-binding motif A (p-loop)
F:598-631/Region: actin binding #status predicted
F:705-727/Region: actin binding #status predicted
F:888-2017/Domain: coiled coil #status predicted <COI>
F:888-1328/Region: S2
F:1329-2017/Region: light meromyosin
F:176/Modified site: N6,N6-trimethyllysine (lys) #status predicted
F:231/Binding site: ATP (lys) #status predicted
F:745,755/Active site: Cys #status predicted

Alignment Scores:
Pred. No.: 1,648-09 Length: 2017
Score: 279.00 Matches: 157
Percent Similarity: 41.30% Conservative: 147
Best Local Similarity: 21.33% Mismatches: 262
Query Match: 8.20% Indels: 170
DB: 1 Gaps: 29

US-09-602-362E-15 (1-2030) x A36014 (1-2017)
QY 81 GACATGCAAACTTTCAAGCAGAGCTCCCGAAGGCAATCTGCCTTCGAGCCTGCCATT 140

```

```

Db 1170 AspLeuGluAlaGluLysAla-----AlaArgAlaLysAlaGlu 1182
QY 141 GAAATGCAAAAGTCTGTTCCTCAAAATAGAGCCCTTGAATTTGAAGAATGAA----- 188
Db 1183 LysValArgAspLeuSerGluGluLeuGluAlaLeuLysAsnGluLeuLysAspSer 1202
QY 189 ---CAACATTGAGAGAGATGAGATATCTCCATCAGATCCAAACAA----- 233
Db 1203 LeuAspThrThrAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1222
QY 234 ---AAGGACTATGAGAAAGTTCTTGGGATCTCGAGAGTCTCTGAGACTGTTTCA 287
Db 1223 LeuLysLysSerLeuGluGluGluThrValAsnHisGluGluValLeuAlaAspMetArg 1242
QY 288 CAGAAGATGTGTTTACCCAGGCTACACATCAAAAGCAAAATAGATAAATAATGGA 347
Db 1243 HisLys-----HisSerGlnGluLeuAsnSerIleAsnAsp 1254
QY 348 AAATTAGAA----- 356
Db 1255 GlnLeuGluAsnLeuArgLysAlaLysThrValLeuGluLysAlaLysGlyThrLeuGlu 1274
QY 357 ---GAGTCTCTGAT-----AAT 371
Db 1275 AlaGluAsnAlaAspLeuAlaThrGluLeuArgSerValAsnSerSerArgGlnGluAsn 1294
QY 372 GATGCTTTTTCGAGGCTCCTCGAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTAGAA 431
Db 1295 AspArgArgLys-----GlnAlaGluSerGlnIleAlaGluLeuGlnValLys 1311
QY 432 TTGATGCACATGCAAACTTTCAAGCAGAGCCCTCCGAGAGCCATCTGCTCTTCGAG--- 488
Db 1312 LeuAlaGluIleGluArgAlaArgSerGluLeuGlnGlnLysCysThrLysLeuGlnGln 1331
QY 489 -----CCTGCCATTGAAATGCAAAAGTCTGTTCCTCAAT 521
Db 1332 GluAlaGluAsnIleThrAsnGlnLeuGluGluAlaGluLeuLysAlaSerAlaAlaVal 1351
QY 522 AAAGCCTTGAATTTGAAGATGAACAAACATTCGAGAGCAGATCAGATGTTCCCTTCAGAA 581
Db 1352 LysSerAlaSerAsnMetGluSerGlnLeuThrGluAlaGlnGlnLeuLeuGluGlu 1371
QY 582 TCAAAACAAAAG-----AAGGTTGAAGAAAATCTTGGGATCTCGAGAGT 626
Db 1372 ThrArgGlnLysLeuGlyLeuSerSerLysLeuArgGlnIleGluSerGluLysGluAla 1391
QY 627 CTCGCTGAGACTTTTCACAGAGGATGTGTGTACCCAGCCTACACATCAAAA--- 683
Db 1392 LeuGlnGlnGlnLeuGluAspGluAla-----LysArgAsnTyrGluArgLys 1409
QY 684 -----GAAATGGATAAAATAGTGGAAAATAGTGGAAAATAGCAATTCACCTAGC 725
Db 1410 LeuAlaGluValThrThrGlnMetGlnGluIleLysLysLysAlaGluGluAspAlaAsp 1429
QY 726 CTATCAAAATCTTGGATCAGTTCATTCTGTGAAAGAGCAGAGGAACTTCAAAAAGAT 785
Db 1430 LeuAlaLysGluLeuGlu-----GluGlyLysLysArgLeuAsnLysAsp 1444
QY 786 CACTGTGAACACACTACAGGAAAATAGGAAACAAATGGAAGAAAGTGTGTGTACTGAAA 845
Db 1445 -----IleGluAlaLeuGluArgGlnValLysGluLeuIleAlaGlnAsn 1459
QY 846 AAGAACTGTCAAGACA---AAGAAATAAATACAGTTAGAGAACCAAAAGTTAA 902
Db 1460 AspArgLeuAspLysSerLysLysIleGlnSerGluLeuGluAlaThrIleGlu 1479
QY 903 TGGGAACAAGACTCTGCAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGA 962
Db 1480 LeuGluAlaGln-----ArgThrLysValLeuGluLeuLysGlnLys 1495
QY 963 AATGCCGAT-----ATATTAAATGAAAAATTAGGGAAGAAATTA 1001

```

Db 1496 AsnPheAspLysIleLeuAlaGluGlySerGluGlnIleAlaGlnGluArg 1515
 QY 1002 GGAAGATCGAA-----GAGCAGCATAGGAAGAGTTAGAACTGAAACAA 1049
 Db 1516 AspThrAlaGluArgGluAlaArgGluGlyThrLysValLeuSerValSerArgGlu 1535
 QY 1050 CTTGAACAGGCTCTC--AGAAATCAAGATATAGAA-----TTGAAGAGTGTAGAACT 1100
 Db 1536 LeuAspGluAlaPheAspLysIleGluAspLeuGluAsnLysArgLysThrLeuGlnAsn 1555
 QY 1101 AATTGGAATCAGTTTCTCACACTCATGAAATGAAATATCTCTTACATGAAATTCG 1160
 Db 1556 GluLeuAspLeuAlaAsnThrGlnGlyThrAlaAspLysAsnValHisGlu----- 1573
 QY 1161 ATGTTGAAAAAGGAATTCCTAAACTGCAATAGCCACACTGAAACACCAATAC 1220
 Db 1574 --LeuGluLysAlaLysArgAlaLeuGluSerGlnLeuAlaGluLeuLysAlaGlnAsn 1592
 QY 1221 CAGAAAAAGGAAT-----AAATCTTTGAGACATTAAGATTTTAAAGAAAGAT 1274
 Db 1593 GluGluLeuGluAspAspLeuGlnLeuThrGluAspAlaLysLeuArgLeuGluValAsn 1612
 QY 1275 -----GCTGAACCTTCAGATGACCTCAAACTGAAAGAGGAATCATTAAC 1319
 Db 1613 MetGlnAlaLeuArgSerGlnPheGluArgAspLeuLeuAlaLysGluGluGlyAlaGlu 1632
 QY 1320 AAAAGGCGATCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAG----- 1367
 Db 1633 GluLysArgArgGlyLeuValLysGlnLeuArgAspLeuGluThrGluLeuAspGluGlu 1652
 QY 1368 -----AACCAATGCTCCTCACTTCTAAATGGAAGGAAACCAAGCAAAAGATA 1415
 Db 1653 ArgLysGlnArgThrAlaAlaValAlaSerLysLysLeuGluGlyAspLeuLysGlu 1672
 QY 1416 CTAGAGCAGAAATGAATCACCACCTCTAGACTGGCTTCTGCTGTACAAGACCATGAT 1475
 Db 1673 IleGluThrThrMetGluMetHisAsnLysValLysGluAspAlaLeu-LysHisAlaLys 1692
 QY 1476 CAATTTGTCATCAAGAAAGGTCAGAACCTGCTTCCACATTCACAGGAGATGCTGT 1535
 Db 1692 slyLeu-----GlnAlaGlnValLysAspAlaLeuArgAspAlaGluGluAlaLysAl 1710
 QY 1536 TTGCAAGAAATGAATGTGTATG-----TGAGTAGTACCGATATATAA 1580
 Db 1710 aAlaLysGluGluLeuGlnAlaLeuSerLysGluAlaAspGlyLysValLysAlaLeuG 1730
 QY 1581 CAATGAGTGCTCCAT-----CAACCATCTTCTGAAAGCTCAAGGAAATCCANAAGCT 1634
 Db 1730 uAlaGluValLeuGlnLeuThrGluAspLeuAlaSerSerGluArgAlaArgAlaAl 1750
 QY 1635 AAAAATTAATCTCAATTATCGAGAGATGCTTAAGAGAAATATACATTCGTTTCAGGAAC 1694
 Db 1750 aGluThrGluArgAspLeuAlaGluGluIleAlaAsnAsnAlaAsnLysGlySerLe 1770
 QY 1695 ATGCACAAAGACCAACGTCAGAACAGTCAATGAAGAGCTGAACACATCTTTC 1754
 Db 1770 uMetIleAsp-GluLysArgArgLeuGluAlaArgIleAlaThrLeuGluGluLeuG 1790
 QY 1755 AAACGAAACAGAT-----NATGTGAACAAACACACTGANCAGCAGGAGCTCTC 1802
 Db 1790 luGluGluGlnSerAsnSerGluValLeuLeuAspArgAlaAlaAlaArgGlnLeuGln 1810
 QY 1803 TAGATCAGAAATTAATTCACACTCAAGCAAAATATGTGGCTTCAACAGCAATTAGTTC 1862
 Db 1810 leGluGlnLeuThrThrGluLeu----- 1817
 QY 1863 ATGCACATTAAGAACTGACACAAACAGCAAGATTAACAATTCATNTCTTGAG 1922
 Db 1818 --AlaAsnGluLysSerAsnSerGlnLys-----A 1827
 QY 1923 AGGAAATATGNCATCATCTTCTTAAAGAGAAAAATGAGGAGATA 1966
 Db 1827 snGluAsnGlyArgAlaLeuLeuGluArgGlnAsnLysGluLeu 1841

RESULT 6

T08880
 NMDA receptor-binding protein yotiao - human
 C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: T08880
 R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
 J. Neurosci. 18, 2017-2027, 1998
 A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
 A:Reference number: Z16511; MUID:98151389; PMID:9482789
 A:Accession: T08880
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1642 <LIN>
 A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AAB86384.1; PID:g2623068
 C:Genetics:
 A:Map position: 7q21-22
 C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Alignment Scores:

Pred. No.:	1-75e-09	Length:	1642
Score:	278.50	Matches:	158
Percent Similarity:	39.52%	Conservative:	172
Best Local Similarity:	18.92%	Mismatches:	250
Query Match:	5.18%	Indels:	255
DB:	2	Gaps:	30

US-09-602-362E-15 (1-2030) x T08880 (1-1642)

QY 135 GCCATTGAAATGCAAAAGCTGTCCAAATAAAGCTTGGAAATCGAAGATCAA----- 188
 Db 112 SerValGluLeuGluSerGluIleSerThrAlaAspAspCysSerSerGluValAsn 131
 QY 189 -----CAAACATTGAGACGATGAGATA 212
 Db 132 GlyCysSerPheValMetArgThrGlyLysProThrAsnLeuLeuArgGluGluPhe 151
 QY 213 CTCCCATGAGATCCAAACAAAGGACTATGAAGAAAGTTCTTGGGATTCGAGAGTCTC 272
 Db 152 GlyValAspSerTyrSer-----GluGlnGlyAlaGlnAspSerProThrHis 168
 QY 273 TGTGACACTGTTTTCACAGAAAGGATGTGTTCACCAAGCTTACACATCAAAAAAGATA 332
 Db 169 LeuGluMetMetGluSerGlu-----LeuAlaGlyLysGlnHisGluIle 183
 QY 333 GATAAAATAATGGAATAATTAGAAGATCTCCTGATAATGATGTTTCTGAAGGCTCCC 392
 Db 184 GluGluLeuAsnArgGluLeuGluGlu----- 192
 QY 393 TGCAGATGAAGATTCTTATTCCAACTAAAGCTTAGAATTTAGGACATCGAACATCTTC 452
 Db 193 -----MetArgValThrTyrGlyThrGluGlyLeuGln-----GlnLeuGlnGluPhe 208
 QY 453 AAAGCAGAGCTCCCGAGAGACCATCTGCTTCGAGAGCTGCCATTGAAATGCAAAAGTCT 512
 Db 209 GluAlaAlaIleLysGlnArg-----AspGlyIleIleThrGlnLeuThr 223
 QY 513 GTTCCAAATAAAGCTTGGAAATTTGAAGATGAACAAACATTGAGACGATGATGATGTC 572
 Db 224 AlaAsnLeuGlnGlnAlaArgArgGluLysAspGluThrMetArgGluPheLeuGluLeu 243
 QY 573 CCTTCAGATCAAAACAAAGAGGCTTGA-----GAAAAATCTTGGGATTCGAGAGTCTC 629
 Db 244 ThrGluGlnSerGlnLysLeuGlnIleGlnPheGlnGlnLeuGlnAlaSerGluThrLeu 263
 QY 630 CTGAGACTGTT-----TCACAGAGGATGTGTGTGATCCCAAG-----GCT 671
 Db 264 ArgAsnSerThrHisSerSerThrAlaAlaAspLeuLeuGlnAlaLysGlnGlnIleLeu 283
 QY 672 ACACATCAAAAGAAATGATAAATAAGTGGAAAAATTAGAAGAT----- 716
 Db 284 ThrHisGlnGlnLeuGluGluGlnAspHisLeuLeuGluAspTyrGlnLysLysLys 303

Qy	717	-----TCAACTAGCCTATCAAAAATCTTGGAT-----ACAGTTCAATCTTGTGAA	761
Db	304	GluAspPheThrMetGlnIleSerPheLeuGlnGluLysIleLysValTyrGluMetGlu	323
Qy	762	AGACCAAGGGAACCTTCAAAAAGATCAC-----TGT	791
Db	324	GlnAspLysLysValGluAsnSerAsnLysGluGluIleGlnGluLysGluThrIleIle	343
Qy	792	GAACACGTCACAGGAAAATCGAACAAATGAAAGAAAGTTTCTGTACTGMAAAGAAA	851
Db	344	GluGluLeuAsnThrLysIleIleGluGluGluLysLysThrLeuGluLeuLysAspLys	363
Qy	852	CTGTCAAGAACGAAAAGAAATAAAATCACAGTTAGAGACCAAAAAGATTAAATCGCAACAA	911
Db	364	LeuThrAlaAspLysLeuLeuGlyGluLeuGlnGluGlnIleValGlnLysAsnGln	393
Qy	912	GAGCTCTGCAGTGTGAGATTGACTTTTAAACCAAGAGAGAGAGAAGAAATCGCCGAT	971
Db	384	GluIleLysAsnMetLysLeuGluLeuThrAsnSerLysGlnLysGluAArgGlnSer---	402
Qy	972	ATATTAAATGAAAAATTAGGCGAAGATTAGCAAGAAATCGAAGAG-----	1016
Db	403	-----SerGluGluIleLysGlnLeuMetGlyThrValGluGluLeuGlnLysArgAsn	420
Qy	1016	-----	1016
Db	421	HisLysAspSerGlnPheGluThrAspIleValGlnArgMetGluGlnGluThrGlnArg	440
Qy	1016	-----	1016
Db	441	LysLeuGluGlnLeuArgAlaGluLeuAspGluMetTyrGlyGlnGlnIleValGlnMet	460
Qy	1017	-----CAGCATPAGAAAGAGTTAGAACTG-----AAA	1043
Db	461	LysGlnGluLeuIleArgGlnHisMetAlaGlnMetGluGluMetLysThrArgHisLys	480
Qy	1044	CAACAACTTGAACGGCTCTCAGA-----	1067
Db	481	GlyGluMetGluAsnAlaLeuAArgSerTyrSerAsnIleThrValAsnGluAspGlnIle	500
Qy	1068	-----ATACAAGATATAGAATTGAAGAGTGTAGAAAGTAATTTGAAT	1109
Db	501	LysLeuMetAsnValAlaIleAsnGluLeuAsnIleLysLeuGlnAspThrAsn---Ser	519
Qy	1110	CAGTTTCTCACACTCATGAAATGAAATATCTCTTACATGAAATTCATGTTGCAA	1169
Db	520	GlnLysGluLysLeuLysGluGluLeuGlyLeuIleLeuGluGluLysCysAlaLeuGln	539
Qy	1170	AAGGAAATGTCATGCTTAAACTGGAATATGCCACACTGAAACACCAATATACCAG-----	1223
Db	540	ArgGlnLeuGluAspLeuValGluLeuSerPheSerArgGluGlnIleGlnArgAla	559
Qy	1224	-----GAAAGGAAAAATAATATAC-----	1241
Db	560	ArgGlnThrIleAlaGluGlnGluSerLysLeuAsnGluAlaHisLysSerLeuSerThr	579
Qy	1242	TTTCAGGACATTAAGATTTTAAAAAGAAAAGAACTGCTCAACTCAGATGACCCATAAATCG	1301
Db	580	ValGluAspLeuLysAlaGluIleValSerAlaSerGluSerArgLysGluLeuGluLeu	599
Qy	1302	AAAGAGAAATCATTAATAAAGGCGCATCTCAATATATAGTGGCAGCTTAAGTTCTGATA	1361
Db	600	LysHisGluAla-----GluValThrAsnTyrLysIleLysLeuGluMetLeuGlu	616
Qy	1362	GCTCAGAACACAATGCTCACTTCTTAAATTGAAGGAAAAACAAGACAAAGAAATA-----	1415
Db	617	LysGluLysAsnAlaValLeuAspArgMetAlaGluSerGlnGluAlaGluLeuGluArg	636
Qy	1416	-----	1430
Db	637	LeuArgThrGlnLeuLeuPheSerHisGluGluGluLeuSerLysLysGluAspLeu	656

Qy	1431	GAATCACACCATCTCTAGACTGGCTTCTGCTGTATACAGACCATGATCAATATTGCATCA	1490
Db	657	GlulGluHisArgIleAsnIleGluLysLeuLysAspAsn-LeuGlyIleHisTyrLy	676
Qy	1491	AGAAAGATCAAGAACCTG-----CTTTCACATTCGAG-----	1524
Db	676	sGlnGlnIleAspGlyLeuGlnAsnGluMetSerGlnLysIleGluThrMetGlnPheGl	696
Qy	1525	-----GAGATGCTTGTTCGAAAGAAAATGAATGTGATGTAGTAGTACCGATATATAA	1580
Db	696	uLysAspAsnLeuIleThrLysGlnAsnGlnLeuIleLeuGluIleSerLysLeuLysAs	716
Qy	1581	CAATGAGTGCTCCATCAACCATCTTCTGAAGCTCAAGAGAAATCCANAAGCCTAAAAAT	1640
Db	716	p-----LeuGlnGlnSerLeuValAsnSerLysSerGluGluMetThrLeuGlnIl	733
Qy	1641	TAAT--CTCAATTATCAGAGAGATGCTCTAAGAGAAAATACATTTGGTTTCAGAACATG	1697
Db	733	eAsnGluLeuGlnLysGluIleGluIleLeuArgGlnGluLysGluLysGlyThr-L	753
Qy	1698	CACAAAGACACCAACGTGAACACAGTGT-----CAAATCAAGG	1736
Db	753	eGluGlnGluValGlnGlnLeuGlnLeuLysThrGluLeuLeuGluLysGlnMetLysG	773
Qy	1737	AGCTGAACCATGTTNCTAAANGCA-----	1762
Db	773	lulysGluAsnAspLeuGlnGluLysPheAlaGlnLeuGluAlaGluAsnSerIleLeuL	793
Qy	1763	-----CAAGATNATGTGAACAAACACACT-----GANCACAGG	1796
Db	793	ysAspGluLysLysThrLeuGluAspMetLeuLysIleHisThrProValSerGlnGluG	813
Qy	1797	AGTCTCTA-----GATCAGA	1811
Db	813	luArgLeuIlePheLeuAspSerIleLysSerLysLysAspSerValTrpGluLysG	833
Qy	1812	AATATTATTCACATAAAGCAAAATATGCGCTTCAACAGCAATTAGTTTCATGCACATA	1871
Db	833	luIleGluIleLeuIleGluGluAsnGluAspLeuLysGlnGlnCysIleGlnLeuAsnG	853
Qy	1872	ANGAAGCTGACACAAACAGATAACAATTTGATNTTCATTNCTTGTAGAGAAAAATG	1931
Db	853	luGluIleGluLysGlnArgAsnThr-PheSerPheAlaGluLysAsnPheGluValAsnT	873
Qy	1932	CNCATCATCTTCTAAAGAG-----	1951
Db	873	yrGlnGluLeuGlnGluGluTyrAlaCysLeuLeuLysValLysAspAspLeuGluAspS	893
Qy	1952	--AAAAATGAGGAGATATTTNATTACNATAACCAATTTAAAA	1990
Db	893	erLysAsnLysGlnGluLeuGluTyrLysSerLysLysLysLys	906

Pred. No.:	2,66e-09	Length:	1690
Score:	275.50	Matches:	174
Percent Similarity:	37.96%	Conservative:	127
Best Local Similarity:	21.94%	Mismatches:	283
Query Match:	8.10%	Indels:	229
DB:	2	Gaps:	35
US-09-602-362E-15 (1-2030) x T13030 (1-1690)			
QY	39	GAAGTGAAGTTCTTATTCCTCAACTAAGCCCTTAGAATTGATGACATG-----CAAACT	92
DB	330	GIYThrMetAsnSerIleAlaThrAlaThrSerArgMetArgMetAsnAlaGlnGln	349
QY	93	TTCAAAGCAGCGCTCCCGAAGAGCCATCTGCGCTCGAGCCCTGCCAATTGAAATG-----	146
DB	350	ArgLysSerSerThrProValLysProIleLeuAlaThrProLysSerGlnPheSerMet	369
QY	146	-----	146
DB	370	GlnAspLeuLeuArgGluLysGlnHisValGluLysLeuMetValArgAspLeu	389
QY	147	---CAAAGTCTCTTCCAAATAAGCCCTTGAATTTG---AAGAAATGAACAAACATTTGAGA	200
DB	390	AspArgGluAspAlaGlnHisGlnAlaLeuGlnLeuGlnLysAsnIleAsnGluLeuLys	409
QY	201	GCA-----GATGAGATCTCCCATCAGAAATCCAAACAAAGAGCATATGAA	245
DB	410	AlaArgIleValGluLeuGluSerAlaLeuGluArgLysLysThrGluGluLeu	429
QY	246	GAAGTCTTGGATCTGAGACTCTCTGT-----GAGACTGTTTCAGAAAGCATGTG	299
DB	430	GlnCysSerIleAspGluAlaGlnPheCysGlyAspGluLeuAsnAlaGlnSerGlnVal	449
QY	300	TGTTTACCAAGGCTACATCAAAAGAAATAGATAAAATAATGGAATA-----	353
DB	450	-----TyrLysGluLysIleHisAspLeuGluSerLysIleThrLys	463
QY	354	-----GAGAGTCTCTGATATGAT	374
DB	464	LeuValSerAlaThrProSerLeuGlnSerIleLeuProProAspLeuProSerAspAsp	483
QY	375	GGTTTCTGAAG-----GCTCCCTGCAGAAATGAAATTTCTTCCAACTAAAGCC	425
DB	484	GlyAlaLeuGlnGluGlurIleAlaGlnLeuGlnGluLysMetThrIleGlnLysGlu	503
QY	426	TTAGAATTGATG-----	437
DB	504	ValGluSerArgIleAlaGluGlnLeuGluGluGlnArgLeuArgGluAsnValLys	523
QY	438	-----GACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAGCCATCTGCCTTC	485
DB	524	TyrLeuAsnGluGlnIleAlaThrLeuGlnSerGluLeuValSerLysAspGluAlaLeu	543
QY	486	GAGCCTGCCAATGCAATGCAAGTCTGTTCCAAATGAAAGCCCTTGAATTG-----	536
DB	544	GluLysPheSerLeuSerGluCysGlyIleGluAsnLeuArgArgGluLeuAlaLeuLeu	563
QY	537	-----AAGAAATGAA-----CAAACTTGAGCAGCATCAGATGTTCCCTTCAGAAATCAAAA	587
DB	564	LysGluGluAsnGluLysGlnAlaGlnGluAlaGlnAlaGluPheThrArgLysLeuAla	583
QY	588	CAAAAGAAGTTGAAGAAATTTCTGGGATTTCTGAG-----	623
DB	584	GluLysSerValGluValLeuArgLeuSerSerGluLeuGlnAsnLeuLysAlaThrSer	603
QY	624	---AGTCTCGT---GAGACTGTTTCACAGAGGATGTGTGTGATCCCAAGGCTACACAT	677
DB	604	AspSerLeuGluSerGluArgValAsnLysSerAspGluCysGluIleLeuGlnThrGlu	623
QY	678	-----CAAAAAAAGATGATAAATAAGTGA	704
DB	624	ValArgMetArgAspGluGlnIleArgGluLeuAsnGlnGlnLeuAspGluValThrThr	643

QY	705	AAATTAGAA-----GATTCAACTAGCTATCAAAATCTTGGATACAGTTCAAT	752
DB	644	GlnLeuAsnValGlnLysAlaAspSerSerAlaLeuAspAspMetLeu-----	659
QY	753	TCTTGTGAAGACCAAGGGAACCTTCAAAAAGATCACTGTGGAACAACAGTACAGGAAATG	812
DB	660	-----ArgLeuGlnLysGluGlyThrGluGluLysSerThrLeuLeu	673
QY	813	GAACAAATGAAAGAAAGATTGTGTACTGAAAGAAACTGTCTAGAACAAAGAAATA	872
DB	674	GluLysThrGluLysGluLeuValGlnIleLysGluGlnAlaLysThrLeuGlnAsp	693
QY	873	AAATCAGATTAGAACCAAAAGTAATGGACACAGACTCTGCAGTGTGAGATTG	932
DB	694	LysGluGlnLeuGluLysGln-----IleSerAspLeuLysGln	706
QY	933	ACTTTAAACCAAGAAAGAGAGAGAAATGCCGATATATTAATGAAATAATAGG	992
DB	707	LeuAlaGluGlnLysLeuValArgGluLysThrGluAsnAlaIleAsnGlnIleGln	726
QY	993	GAGAATTTAGGAAGATCGAAGAGAG-----CATAGGAAGAGCTAGAA-----	1037
DB	727	LeuGluLysGluSerIleGluGlnGlnLeuAlaLeuLysGlnAsnGluLeuGluAspPhe	746
QY	1038	GTGAACAACAACCTTGAAACAGGCTCTCAGAATAACAAGATATAGAA-----	1082
DB	747	GlnLysLysGlnSerGluSerGluValHisLeuGlnGluIleLysAlaGlnAsnThrGln	766
QY	1083	-----TTGAAGAGTGTAGAAAGTAATTTGAAT	1109
DB	767	LysAspLeuGluLeuValGluSerGlyGluSerLeuLysLeuGlnGlnLeuGlu	786
QY	1110	CAGTTTCTCACACTCATGAAATGAAATTTATCTCTTACATGAAATTTGCATTTGAAA	1169
DB	787	GluLysThrLeuGlyHisGluLysLeuGlnAlaAlaLeuGlu-----LeuLys	803
QY	1170	AAGGAA-----ATTGCCATCTGAAACTGAAATAGCCACACTGAAACACCAA	1217
DB	804	LysGluLysGluThrIleLysGluLysGluGlnGluLeuGlnGlnLeuGlnSerLys	823
QY	1218	TACCAGAAAGGAAAT-----	1235
DB	824	SerAlaGluSerGluSerAlaLeuLysValValGlnValGlnLeuGluGlnGlnGln	843
QY	1236	-----AAATCTTTGAGGACATT	1253
DB	844	GlnAlaAlaAlaSerGlyGluGluGlySerLysThrValAlaLysLeuHisAspGluIle	863
QY	1254	AAGATTTTAAAGAAAGAAATGCTGAACTTCAGATGACCTTAAACTGAAAGAGGAAATCA	1313
DB	864	SerGlnLeuLysSerGlnAlaGluGluThrGlnSerGluLeuLysSerThrGluSerAsn	883
QY	1314	TTAACTAAAGGCGATCTCAATAT-----AGTGGCGAGCTTAAAGTTCTTGATAGCT	1364
DB	884	LeuGluAlaLysSerLysGlnLeuGluAlaAlaAsnGlySerLeuGlu-----	899
QY	1365	GAGAACCAATGTCTCTCTTAATTTGAAGAAAAACAGACAAAGAAATCTAGAGGCA	1424
DB	900	GluGluAlaLysLysSerGlyGlnLeuGlnGluGlnIleThrLys-----LeuLysSer	917
QY	1425	GAAATTCATCACACCATCTCAGCTGCTGCTGTGTACAGACCATCATCAATTTGTG	1484
DB	918	GluValGluGluThrGlnAlaAlaLeuSerSerTyr-----HisThrAspValGlu	934
QY	1485	ACATCAAGAAAAAGTCAAGAACCTGCTTTCCATTCAGGAGATGCTGTTTTCGCAAGA	1544
DB	935	SerLysThrLysGlnLeuGlu-----AlaAlaAsnAlaAlaLeuGlu-----	948
QY	1545	AAATGATGATTTGAT-----GTGAGTAGTAGCCGATATATAACAAATGAGGT	1589
DB	949	LysValAsnLysGluTyrAlaGluSerArgAlaGluAlaSerAspLeu-GlnAspLysVa	968
QY	1590	G-----CTCCATCAACCACTTTCTTGAAGCTCAAGGAAATCCANAAG	1631

Db 968 llysgluLeuThrAspThrLeuHisAlaGluLeuAlaGluArgSerSerSerAl 988
QY 1632 CTTAAATTAATCTCAATATGCGAGCATGCTCTAAGAGAAATACAT 1681
Db 989 aleuHisThrLysLeuSerLysPheSerAspGluLeuAlaThrGly-HisLysGluLeu 1008
QY 1682 -----TGGTTTCAGGAACATGCGCAAAAGACCAACGCTGAACACAGT 1724
Db 1008 hrSerLysAlaAspAlaTrpSerGlnGluMetLeuGlnLysGluLys 1023
QY 1725 GTCAATGAGGAAGCTGACACATGTNTCAANGCAACAGATNATGTGACAAACACA 1784
Db 1024 --GluLeuGlnGluLeuArgGlnGlnLeuGlnAspSerGlnAspSerGlnThrLysLeu 1043
QY 1785 CTGANCAGCAGGAGTCTCTAGATCAGAAATTTTCAACTACAAAGCAAAATATGTGGC 1844
Db 1043 ysAlaGluGlyGluArg--LysGluLysSerPheGluGluSerIleLysAsn 1060
QY 1845 TTCAACAGCAATAGTTTCATGCGACATAANGAAAGC 1879
Db 1060 euGlnGluGluValThrLysAlaLysThrGluAsn 1071

RESULT 8
C71622
hypothetical protein PF0145c - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
R/Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
S.; Partea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1133, 1998
A/Title: Chromosome 2, sequence of the human malaria parasite Plasmodium falciparum.
A/Reference number: A71600; MUID:99021743; PMID:9804551
A/Accession: C71622
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-1979 <GAR>
A/Cross-references: GB:AE001375; GB:AE001362; NID:93845105; PIDN:AC71819.1; PID:9384510
A/Experimental source: clone 3D7
C/Genetics:
A/Gene: PF0145c

Alignment Scores:
Pred. No.: 8,66e-09 Length: 1979
Scores: 267.00 Matches: 185
Percent Similarity: 37.06% Conservative: 130
Best Local Similarity: 21.76% Mismatches: 268
Query Match: 7.85% Indels: 268
Gaps: 38

US-09-602-362E-15 (1-2030) x C71622 (1-1979)

QY 141 GAATGCAAAAGTCGTGTTCCAAATAAGCCCTGGAATGAAGATGAA-----CAAAACA 194
Db 118 GlutLeuSerLysAsnIleHisLysAspAsnGluLeuGluAsnGlnLeuLysAspThr 137
QY 195 TTGAGA-----GCAGATGAGATCTCCCATCAGAAATCCAA----- 230
Db 138 LeuLysSerIleSerSerLeuSerAsnLysIleValAsnThrGluSerLysIleGluGlu 157
QY 231 -----CAAAAGGACTATGAAGAA 248
Db 158 LeuGluLysGluLeuLysValLysAspLysAsnIleAspAsnAspTyrGluAsn 177
QY 249 AGTCTTGGGATCTGAGAGTCTCTGTCAGACTGTTTCACAGAG-----GATGTGTGTTA 305
Db 178 LysLeuLysGluLysGlu-----AspPheValLysGlnLysIleAspMetLeuAsn 194
QY 306 CCCAAGGCTACA-----CATCAAAAGAAATAGAT-----AAAATA 341
Db 195 GluLysGluAsnLeuLeuGlnGluLysGluLeuAspIleAsnLysArgGluLysLysIle 214

QY 342 AATGGAATA----- 350
Db 215 AsnGluLysGluLysAsnIleIleLysLysGluGluThrPheHisAsnIleGluLysGlu 234
QY 351 ---TTAGAGAGTCTCCTGATAATGATGATGTTTCTGAAGGCTCCTGCAGATGAAGTT 407
Db 235 TyrLeuGluLysAsnLysGluArgGlu----- 243
QY 408 TCTATTCCAACTAAAGCCCTTAGAATTCATGAC-----ATGCAAACTTCAAA 455
Db 244 -----ThrIleSerIleGluIleIleAspIleLysLysHisLeuGluLysLeuLys 260
QY 456 GCAGAGCTCCCGAGAGCCATCTGCTTCGAGCCTGCCATTCGAATGCAAAATCTGTGTT 515
Db 261 IleGluIleLysGluLysLysGluAspLeuGluAsnLeuAsnLysLysLeuLeuSerLys 280
QY 516 CCAAAATAAGCCCTTGGAAATGAAG-----AATGAACAACATTCAGAGCA 560
Db 281 GluAsnValLeuLysGluLeuLysGlyCysValLysGluLysAsnGluThrIleAsnSer 300
QY 561 -----GATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAA----- 602
Db 301 LeuAsnAspAsnIleIleGluLysGluLysLysTyrLysLeuLeuGluLeuGlu 320
QY 603 -----GAAATTCCTGGGATTCAGAGT 626
Db 321 GluLysAsnLysGlnIleAspLeuLeuAsnLysGlnGluLysGluLysGluLys 340
QY 627 CTCGCTGAGACTCTTTCACAGAGAGGATGTGTGTACCCCAAGCTACACATCAAAAGAA 686
Db 341 GluArgGluLysGluLysGluArgGlu-----LysGluLysGluLysGlu 355
QY 687 ATCGATAAATAAGTGGAAATTAGAAGATCAACTAGCCTATCAAAAATCTTGGATACA 746
Db 356 TyrAspThrLeuIleLysGluLeuLysAspGluLys-----IleSerIleLeuGluLys 373
QY 747 GTTCATCTCTGTGAA---AGAGCAAGGGAACTCAAAACAT-----CAC 788
Db 374 ValHisSerIleLysValArgGluMetAspIleGluLysArgGluHisAsnPheLeuHis 393
QY 789 TGTGAACAACGTCACAGAAATGGAACAAATGAAAGAAAGATTTTGT----- 836
Db 394 Met-GluAsp-----GlnLeuLysAspLeuLysAsnSerPheValLysAsnAsn 410
QY 837 GTACTGAAAGAAAGAACTGTACAGAACAAAGAAATAAATCACTAGTAGAACCAAAA 896
Db 411 GlnLeuLysValTyrLysCysGluIleLysAsnLeuLysThrGluLeuGlu----- 427
QY 897 GTTAAATGGCAACAACAGAGCTCTGC-----AGTGTGAGATTG 932
Db 428 ---LysLysGluLysGluLeuLysAspIleGluAsnValSerLysGluLeuLeuAsnLys 446
QY 933 ACTTTAAACCAAGAGAGAGAGAGAAATGCCGATATATTAATGAAATAATTAGG 992
Db 447 LeuIleAsnGlnLeuAsnGluLysGluLysGlnIleLeuAlaPheAsnLysAsnHisLys 466
QY 993 GAAGAA-----TTAGGAAGATCGAA 1013
Db 467 GluGluIleHisGlyLeuLysGluGluLeuLysGluSerValLysIleThrLysIleGlu 486
QY 1014 GAGCAGCATAGGAAGAGTTA---GAAGTGAACAA-----CAACTTGAACAG 1058
Db 487 ThrGlnGluLeuGlnGluMetValAspIleLysGlnLysGluLeuAspGlnLeuGlnGlu 506
QY 1059 GCTCTCAGAAATACA-----GATATGAATTCGAAGAGTGAAGAGTAATTG 1106
Db 507 LysTyrAsnAlaGlnIleGluSerIleSerIleGluLeuSerLysLysGluLysGlu 526
QY 1107 AATCAGGTTCTCACACTCATGAAAATGAA----- 1136
Db 527 AsnGlnTyrAsnThrThrIleGluIleAsnAsnLeuAsnGluLysLeuGluGlu 546
QY 1137 -----AATTATCTCTTA 1148

Db	547	ThrAsnLysGluTyrThrAsnLeuGlnAsnAsnTyrThrAsnGluIleAsnMetLeuAsn	566	Db	892	GluGlnIleLeuLeuThrGluMetGluGluLeuLysCysGlnArgAspAsnLysTyrSer	911
Qy	1149	CATGAAATTCATGTTGAAAGGAAATTCATGCTAAACATGGAATAGCCACACTG	1208	Qy	2003	TCAATATGGAATAAAAAAAAAAAAAA	2026
Db	567	AsnAspIleHisMetLeuAsnGlyAsnIleLysThrMetAsnThrGlnIleSerThrLeu	586	Db	912	AspLeuTyrGluLysTyrIleLys	919
Qy	1209	AAA-----CACCAATACCAGGAAGGAAAAATAAATACTTTGAGGACATTAAAGATT	1259	RESULT 9			
Db	587	LysAsnAspValHisLeuLeuAsnGluGlnIleAspLysLeuAsnAsnGluLysGlyThr	606	T14156			
Qy	1260	TTAAAGAAAAGATGCTCAACTTAC-----ATGACCTTAAACTGAAGAGGAA	1310	kinesin-related protein - African clawed frog			
Db	607	LeuAsnSerLysIleSerGluLeuAsnValGlnIleMetAspLeuLysGluGluLysAsp	626	C;Species: Xenopus laevis (African clawed frog)			
Qy	1311	TCATTAACTAAAGGCGCATCTCAATATATGCGCGAGCTTAAAGTTCTGATA-----	1361	C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000			
Db	627	PheLeuAsnAsnGlnIleValAspLeuSerAsnGlnIleAspLeuLeuThrArgLysMet	646	C;Accession: T14156			
Qy	1362	-----GCTGAGAACAAATGCTCTTCTAAATTTGAAGGAAAAACAAGACAAAGATA	1415	R;Wood: K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.			
Db	647	GluGluLysGluAsnLysMetLeuGluGlnGluAsnLysTyrLysGlnGluMetGluLeu	666	Cell 91, 357-366, 1997			
Qy	1416	CTAGAGGCAGAAATT-----	1430	A;Title: CENP-B is a plus end-directed kinetochore motor required for metaphase chromosome			
Db	667	LeuArgGlyAsnIleLysSerSerGluAsnIleLeuAsnAsnAspGluValCysAsp	686	A;Reference number: Z17893; MUID:98028574; PMID:9363944			
Qy	1431	-----GATCACCACCAT	1442	A;Accession: T14156			
Db	687	LeuLysArgLysLeuSerLeuLysGluSerGluMetLysMetMetLysGluGluHisAsp	706	A;Molecule type: mRNA			
Qy	1443	CCTAGACTGGCTCTGCTGTACAGACCATGATCAATTTGTGACATCAA-----	1491	A;Residues: 1-2954 <WOO>			
Db	707	LysLysLeuAlaGluLeuLysAspCysAspValArgIle-ArgGluMetAsnGluLys	726	A;Cross-references: EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AAC60300.1			
Qy	1492	-----GAAAAAGTCAAGAACCTGCTTTCCACATTGCA	1523	C;Genetics:			
Db	726	sAsnGluAspLysIleAsnMetLeuLysGluGluTyrGluAspLysIleAsnThrLeuLys	746	C;Superfamily: centromere protein E; kinesin motor domain homology			
Qy	1524	GGAG-----ATGCTGTTTGGCAAGAAAATCAATGTTTCATGTGAGTAGT	1568	Alignment Scores:			
Db	746	sGluGlnAsnGluAspLysIleAsnThrLeuLysGluGln-----	759	Pred. No.: 8,75e-09 Length: 2954			
Qy	1569	ACCGATATATAACATGAGGTGCTCCATCAACCATCTTTCTGAAGCTCAAGGAAATCCAN	1628	Score: 267.00 Matches: 108			
Db	760	-----AsnGluAspLysIleAsnThrLeuLysGluGluTyrGluHisLysIleAs	776	Percent Similarity: 44.00% Conservative: 112			
Qy	1629	AGCCTAAATAATCTCAATATGCGAGAGATGCTCTAAGAGAAATACATTGGTTTC	1688	Best Local Similarity: 21.60% Mismatches: 180			
Db	776	nThrMetLysGluGluTyrGluHisLysIleAsnThrLeuAsnGluGln-AsnGluHisL	796	Query Match: 7.85% Indels: 100			
Qy	1689	AGGAACATGCACAAAGACCAACGAGTGAACACAGTGTCAA---ATGAAGGAA-----	1739	DB: 2 Gaps: 15			
Db	796	ysIleAsnThrLeuAsnGluGlnAsnGluHisLysIleAsnThrMetLysGluGluTyrG	816	US-09-602-362B-15 (1-2030) x T14156 (1-2954)			
Qy	1740	CTGAACACATGNTTCAAAAGCAACAGATATGTGAAACACACACTGANCAGCAGAGT	1799	Qy 12 AAAGATCGTCTTCTGAAGCTAACTCGGAATGAAAGTTTCTATTCCAACTAAAGCTTA	71		
Db	816	luAspLysMetAsnThrLeuAsnGluGlnAsnGluAspLysMetAsnSerLeuLysGluG	836	Db 1874 ArgAspAspLeuAsnGlnSerLysGlnHisLeuPheSerGluIleGluThrLeuSerLeu	1893		
Qy	1800	CTCTAGATCAGAAATTTTCACTCAACAAAGCAAAATATATGCTTCAACAGCAATTAG	1859	Qy 72 GAATTGATGAGCATGCAAACTTTCAAAGCAGACGCTCCCGAG-----AAGCATCTGCC	125		
Db	836	luTyrGluAsnLysIleAsnGlnIleAsnSerAsnAsnGluIleLysIleLysAspValV	856	Db 1894 SerLeuLysGluLysGlu---PheAlaLeuGluGlnAlaGluLysAspLysAlaAspAla	1912		
Qy	1860	TTTCATGCACATAAAGAGCTGCAACAAAGCAAGATTAACATTAATGATNTTCTT	1919	Qy 126 TTGAGCCTCCCATTTGAATGCAAAAGTCTGTTCCAAATTAAGCCTTGGAAATGAGAAT	185		
Db	856	AlaAsnGluTyrIleGluGlu-ValAspLysLeuLysValThrLeuAsp-----	871	Db 1913 AlaArgLysThrIleAspIleThrGluLysIleSerAsnIleGluGlnLeuGln	1932		
Qy	1920	GAGAGAAAT-----GKNCATCATCTTCTAAAGAGAA	1954	Qy 186 GAACAAACATGAGAGCAGATGAGATACCTCCCATCAGAATCCAAACAAAGAGACTATGA	245		
Db	872	GluLysLysGlnPheAspLysGluIleAsnTyrAlaHisIleLysAlaHisGluLys	891	Db 1933 GlnAlaThrAsnLeuLysGluThrLeuTyrGluArgGluSerLeuIleGlnCysLysGlu	1952		
Qy	1955	AATGAGGAGATATTNATTACNATAACCATTTAAAA-----AACCCGATATT	2002	Qy 246 GAAAGTTCTTGGATTTCTGAGTCTCTGTGAGACTGTTTCACAGAGATGTTCTTTA	305		
				Db 1953 GlnLeuAlaLeuAsnThrGluHisLeuArgGluThrLeuLysSerLysAspLeuAlaLeu	1972		
				Qy 306 CCCAGGCTACACATCAAAAGAA-----ATAGATAAA	338		
				Db 1973 GlyLysMetGluGlnGluArgAspGluAlaAlaAsnLysValIleAlaLeuThrGluLys	1992		
				Qy 339 ATAAATCGAAAAATTAGAAGAGTCTCTGATATATGATGTTTCTGAAGGCTCCCTCGAGA	398		
				Db 1993 MetSerSerLeuGluGluGlnIleAsnGluAsnValThrThrLeuLysGluGlyGlu	2012		
				Qy 399 ATGAAAGTTCTTATCCAACTAAAGCCTTAGAATTGATGGACATGCAAACTTTCAAAGCA	458		
				Db 2013 GluLys-----GluThrPheTyrLeu	2019		
				Qy 459 GAGCTCCCGAGAGCCCATCTGCTTCGACCTGCCATTTGAAATGCAAAAGTCTGTTC	518		
				Db 2020 GlnArgProSerLysGlnGlnSerSerGlnMetGluGluLeuArgGluSerLeuLys	2039		
				Qy 519 AATTAAGCCCTTGGAAATTAAGAAATGAACAAACATTGAGACAGATCAGATGTTCCCTTCA	578		

```
Db      2040 ThrLysAspLeuGlnLeu----- 2045
QY      579 GAATCAAAACAAAGAGGTTGAAGAAAATTTCTGGATTCTGAG----- 623
Db      2046 GluGluAlaGluLysGluLeuSerGluAlaThrAsnGluLeuLeuThrAlaLys 2065
QY      624 -----AGTCTCGTGAGACTTTTCACAGAGGATGTGTGTACCAAGGCTACACAT 677
Db      2066 IleSerSerLeuGluGluLeuGluLeuGlnAsnAlaSerIleLeuAsnGluAlaValSer 2085
QY      678 CAAAAGAAATGGATAAATAATAGTGAATAATAGAAATTCACCTATCAAAATC 737
Db      2086 GluArgGluAsnLeuArgHisSer---LysGlnGlnLeuValSerGluLeuGluGlnLeu 2104
QY      738 TTGGATACAGTTCATCTCTGTGAAGAGCAAGGAACTTCAAAAGATCACTGTGAACA 797
Db      2105 SerLeuThrLysSerArgAspHisAlaPheAlaGlnSerLysArgGluLysAspGlu 2124
QY      798 CGTACAGGAAAATGGAACAAATGAAAGAAAGATTTTGTGTACTGAAAGAAATGTCAC 857
Db      2125 AlaValAsnLysIleAlaSerLeuAlaGluGluIleLysIleLeuThrLysGluMetAsp 2144
QY      858 GAAGCAAAAGAAATAAATCAAGTTAGAGAACCAAAAGTTAAATGGAAACAAGAGCTC 917
Db      2145 GluPheArgAspSerLysGluSerLeuGlnGlnSerSerHisLeuSerGluGluLeu 2164
QY      918 TGCAGTGTGAGATTGACTTTTAAACCAAGAGAGAGAGAGAGAAATGCCGATATATTA 977
Db      2165 CysThrTyrrLysThrGluLeuGlnMetLeuLysGlnGlnLysGluAsp-----Ile 2181
QY      978 AATGAAAATATAGGGAAGAAATAGAGAAATCGAAGAG-----CAGCATAGGAAAGAG 1031
Db      2182 AsnAsnLysLeuAlaGluLysValLysGluValAspGluLeuGlnHis----- 2198
QY      1032 TTAGAGTGAACAACAACCTTGAACAGGCTCTCAGAAATACAGATATAGAAATTGAAGAGT 1091
Db      2199 -----LeuSerSer 2201
QY      1092 GTAGAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATATTATCTCTACAT 1151
Db      2202 LeuLysGluGlnLeuAspGlnIleGlnMetGluLeuArgAsnGluLysLeuArgAsnTyrr 2221
QY      1152 GAAATATGTC-----ATGTTGAAAAGGAAATGCCATGCTAAAACCTGGAATA 1199
Db      2222 GluLeuCysGluLysMetAspIleMetGluLysGluIleSerValLeuArgLeu----- 2239
QY      1200 GCCACACTGAAACACCAATACACAGAAAGGAAATAAATACTTTTGAGCATTAAGATT 1259
Db      2240 -----MetGlnAsnGluProGlnGlnGluLysAspValAlaGluArgMetAspIle 2257
QY      1260 TTAAGGAAAGAAAGATGCTGAATTCAGATGACCCCTAAACCTGAAAGAGCAATCACTACT 1319
Db      2258 LeuGluSerArgAsnGlnGluIleGln-----GluLeuMetGlu 2270
QY      1320 TTAAGGCAATCAATATAGTGGGAGCTTAAGTTCTGATAGCTGAGAAACACATGCTC 1379
Db      2271 LysIleSerAlaValTyrrSerGluGln-----HisThrLeuLeu 2283
QY      1380 ACTTCTAAATTCAGGAAACAAACAGACAAAGAAATACTAGAGGAGAAATTCGAATCACAC 1439
Db      2284 SerSerLeuSerSerGlu-----LeuGlnLysGluThrGluAlaHis 2297

RESULT 10
S28261
centromere protein E - human
N:Alternate names: centromere 312K protein; kinesin-related protein CENP-E
C:Species: Homo sapiens (man)
C>Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 19-Jan-2001
C:Accession: S28261
R:ifen, T.J.; Li, G.; Schaar, B.T.; Szilak, I.; Cleveland, D.W.
Nature 359, 536-539, 1992
A:Title: CENP-E is a putative kinetochore motor that accumulates just before mitosis.
A:Reference number: S28261; MUID:93024922; PMID:1406971
```

```
A:Accession: S28261
A:Molecule type: mRNA
A:Residues: 1-2663 <YEN>
A:Cross-references: EMBL:Z15005; NID:g29864; PIDN:CAA78727.1; PID:g29865
C:Genetics:
A:Gene: GDB:CENPE
A:Cross-references: GDB:361164; OMIM:117143
A:Map position: 4q24-4q25
C:Superfamily: centromere protein E; kinesin motor domain homology
C:Keywords: ATP; coiled coil; microtubule binding; mitosis; nucleotide binding; P-loop
P:7-335/Domain; kinesin motor domain homology <KNOT>
P:86-93/Region; nucleotide-binding motif A [P-loop]
P:486-2183/Domain; coiled coil #status predicted <COI>
P:192/Binding site: ATP (Lys) #status predicted

Alignment Scores:
Pred. No.: 9.35e-09 Length: 2663
Score: 266.50 Matches: 172
Percent Similarity: 38.30% Conservative: 126
Best Local Similarity: 22.11% Mismatches: 305
Query Match: 7.83% Indels: 175
DB: 30
```

US-09-602-362E-15 (1-2030) x S28261 (1-2663)

```
QY      12 AAAGATGCTCTCTGAAGGCTAACTGCGAATGAAAGTTTCTTATCCAACTAAAGCCTTA 71
Db      1111 LysGluGlyGluLeuSerArgThrCysAspArgLeuAlaGluGluLysLeuLys 1130
QY      72 GAATGTGATGACATGCAAACTTTCAAGCAGAGAGCTCCCGAAGAGCCATCTGCCTTCGAG 131
Db      1131 Glu-----LysSerGlnGlnLeuGlnGlyGlnGlnGlnLeuLeuAsnValGlnGlu 1148
QY      132 CCTGCCATTGAATGCAAAAG-----TCTGTTCCAAATAAAGCCTTGAATTCGAG 182
Db      1149 GluMetSerGluMetGlnLysLysIleAsnGluIleGluAsnLeuLysAsnGluLys 1168
QY      183 AAT---GAACAAACATTTGACAGCAGATGATATCTCCATCAGAATCCAAACAAAG--- 236
Db      1169 AsnLysGluLeuThrLeuGluHisMetGluThrGluArgLeuGluAlaGlnLysLeu 1188
QY      237 -----GACTATGAAGAAAGTTCTTGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAG 290
Db      1189 AsnGluAsnTyrrGluGlu-----ValLysSerIle-----ThrLysGluArg 1202
QY      291 AAGGATGCTGTGTTTACCACAAAGGCTACACATCAAAAGAAATAGATAAAATAAATGGAATA 350
Db      1203 LysValLeuLysGlnLeuGlnLysSerPheGluThrGluArgAspHisLeuArgGlyTyrr 1222
QY      351 TTAGAAGAGTCTCTGTGATAAATGATGTTTCTGAAGCTCCCTGCAGAAATGAAAGTTTCT 410
Db      1223 IleArgGlu----- 1225
QY      411 ATTCCAACCTAAAGCCTTAGAATTTGATGGACATGCAACATTTCAAAGCAGAGCCTCCGAG 470
Db      1226 -----IleGluAlaThrGlyLeuGlnThrLysGluGluLeuLysIleAlaHisIleHis 1243
QY      471 AAGCCATCTCCCTTCGAGCCTGCCATTGAAATCAAAAGTCTGTTCCAAATAAAGCCTTG 530
Db      1244 LeuLysGluHisGlnGlnThrIleAspGluLeuArgArgSerValSerGluLysThrAla 1263
QY      531 GAATTGAGATGAACAAACATTG----- 554
Db      1264 GlnIleIleAsnThrGlnAspLeuGluLysSerHisThrLysLeuGlnGluLeuPro 1283
QY      555 -----AGAGCAGATCAGATGTTCTCCCTTCAGAATCAAAACAAAGAGGTTGAAGAA 605
Db      1284 ValLeuHisGluGlnGlnGluLeuLeuProAsnValLysLysValSerGluThrGlnGlu 1303
QY      606 AATCTTTGGGATTTCTGAGACTTCCGTGAGACTGTTTTCACAGAAAGATGTGTGTATCCC 665
Db      1304 ThrMetAsnGluLeuGluLeuLeuGlnSerThrThrLysAspSerThrThrLeu 1323
```

666	QY	AGGCTACACATCAAAAAGAATG-----GATAAATAAGTCGAAAAATTAGAAAT	716
		::: ::::	
1324	Dd	AlaArgilegluMetGluArgLeuArgLeuAsnGluysPheGlnGluSerGlnGluGlu	1343
		::: ::::	
717	QY	TCAACTAGCCTATCAAAAATCTTGGATACAGTTTCATTCTTGCGAAAAGACAAGGAAC	776
		::: ::::	
1344	Dd	IleLysSerLeuthrLysGluArgAspAsnLeuysThrIleLysGluAlaLeuGluVal	1363
		::: ::::	
777	QY	CXA-----AAAGATCACTGTGAACACGTCACGAAATAATGGAACAAATGAA	824
		::: ::::	
1364	Dd	LysHisaspGlnLeuLysGluHisIleArgGluThrLeuAlaLysIleGlnGluSerGln	1383
		::: ::::	
825	QY	ACAAG-----TTTTGTGTACTGAAAAGAACTGTCAGAACCAAAGAAATAAAA	875
		::: ::::	
1384	Dd	SerLysGlnGluGlnSerLeuAsnMetLysGluLysAspAsnGluThrThrylIleVal	1403
		::: ::::	
876	QY	TCACAGTTAGAACCAAAAA-----GTTAAATCGGAACAA	911
		::: ::::	
1404	Dd	SerGluMetGluGlnPheLysProLysAspSerAlaLeuLeuArgIleGluIleGluMet	1423
		::: ::::	
912	QY	GAGCTCTGCAGTGTGAGATTGACTTTAAACCAGAAGACAGAG--	956
		::: ::::	
1424	Dd	LeuGlyLeuSerLysArgLeuGlnGluSerHisAspGluMetLysSerValAlaLysGlu	1443
		::: ::::	
957	QY	-----AGAAAGAAATCCCATATATATAAATGAA	983
		::: ::::	
1444	Dd	LysAspAspLeuGlnArgLeuGlnGluValLeuGlnSerCluSerAspGlnLeuLysGlu	1463
		::: ::::	
984	QY	AAAATTAGGGAAGAAATTAGGAAGA-----	1007
		::: ::::	
1464	Dd	AsnIleLysGluIleValAlaLysHisLeuGluThrGluGluLeuLysValAlaHis	1483
		::: ::::	
1008	QY	-----ATCGAAGACAGAGCAT-----AGGAAA	1028
		::: ::::	
1484	Dd	CysCysLeuLysGluGlnGluGluThrIleAsnGluLeuArgValAsnLeuSerGluLys	1503
		::: ::::	
1029	QY	GAGTTAGAA-----GTGAACAACAACACTTTGAACAG-----GTCCTC	1064
		::: ::::	
1504	Dd	GluThrGluIleSerThrIleGlnLysGlnLeuGluAlaIleAsnAspLysLeuGlnAsn	1523
		::: ::::	
1065	QY	AGAATCACAGATATAGAAATTCAGAGGTGAGAAGTAATTGTAATCAGGTTCTCACACT	1124
		::: ::::	
1524	Dd	LysIleGlnGluIleTyrgluLysGluGluLeuAsnIleLysGlnIleSerGluVal	1543
		::: ::::	
1125	QY	CATGAAAATGAAAATTATCTCTTACATGAAAATTGCATGTTGAAAAGGAAATGCCATG	1184
		::: ::::	
1544	Dd	GlnGluAsnValAsn-----GluLeuLysGln	1552
		::: ::::	
1185	QY	CTAAACTGGAANTAGCCACACTGAAACACCAATACCAGAAAAGGAAATAATACTTT	1244
		::: ::::	
1553	Dd	PheLysGluHisArgLysAlaLysAspSerAlaLeuGlnSerIleGluSerLysMetLeu	1572
		::: ::::	
1245	QY	GAGGACATTAAAGATTTTAAAGAAAAAGAAATGCTGAACCTTCAGATGACCCCTAAAACCTGAA	1304
		::: ::::	
1573	Dd	GluLeuThrAsnArgLeuGlnGluSerGlnGluGluIleGlnIleMethIleLysGluLys	1592
		::: ::::	
1305	QY	GAGGAATCATTAATAAAGGGCATCTCAATATAGTGGGCGAGCTTAAAGTTCTGATAGCT	1364
		::: ::::	
1593	Dd	GluGlu-----MethLysArgValGlnGluAlaLeuGlnIleGluArgAspGlnLeuLys	1610
		::: ::::	
1365	QY	GAGAACACATG---CTCACTCTTCAATTCAAGGAAAAACAACAGACAAAGAAATACTAGAG	1421
		::: ::::	
1611	Dd	GluAsnThrLysGluIleValAlaLysMethLysGluSerGlnGluLysGlu-Tyr-----	1628
		::: ::::	
1422	QY	GCAGAAATTGAATCACACCATCTTAGACTCGCTTCTGCTGTACAGACCATGATCAAAATT	1481
		::: ::::	
1629	Dd	-GlnPheLeuLysMetThrAlaValAsn-----GluThrGlnGluLysMe	1643
		::: ::::	
1482	QY	GTGACATCAGAAAAAGTCAGAACCTGCTTTCCACATTCACGAGAGATGCTGTTGCA	1541
		::: ::::	
1643	Dd	cysGluIleGluHisLeuLysGluGlnPheGluThrGlnLysLeuAsnLeuGluAsnII	1663
		::: ::::	
1542	QY	AGAAAAATGAATGTTGATGTGAGTAGTACCGATATATAACAATCAGGTGCTCCATCAACC	1601

Db 1663 eGlu- ||| |||-ThrGluAsnIleArgLeuThrGlnIleLeuHisGlnAs 1677

QY 1602 ACTTCTCGAAGCTCAAAGGAATCCANAAAGCCTAAAAATTAA 1643

Db 1677 nLeuGluGluMetArgSerValThrLysGluArgAspLeuArgSerValGluGluTh 1697

QY 1644 TCTCAATTATGCAGGAGATGCTCTTAACAGAAAATACTGGTTTCAGGACATGCACAAA 1703

Db 1697 rLeuLysValGluArgAspGlnLeuLysGluAsn-----LeuArgGluThrIleThr-A 1715

QY 1704 GAGAC-----CAAGGTGAACACAG--TGTCAAATGAAGGAGCTGAACACATGT 1751

Db 1715 rgaspLeuGluLysGlnGluGluLeuLysIleValHisMetHisLeuLysGluHisGlnG 1735

QY 1752 NTCAAANCGCAACA-----GATNATGTGAACAAACACACTGANACGC 1793

Db 1735 luThrIleAspLysLeuArgGlyIleValSerGluLysThrAsnGluIleSerAsnMetG 1755

QY 1794 AGAGCTCTCTAGATCAGAATATTTCACACTCAAGCAAAAATATGTGCTTCAACAGC 1853

Db 1755 lnyAspLeuGluHisSerAsnAspAlaLeuLysSalGlnAspLeuLysIleGlnGluG 1775

QY 1854 AATTA-----GPTCATGCACATAANGAAGCTGACAAACAAAGCAAGTAACAATTGAT 1907

Db 1775 luLeuArgIleAlaHisMetHisLeuLysGluGlnGlnGluThrIleAsp----- 1791

QY 1908 NTTCAATNTCTGAGAGAAAATGCCATCATCTTCTAAAGAGGAAAATGAGGAGATAT 1967

Db 1792 -----LysLeuArgGlyIleValSerGluLysThrAspLysLeuS 1805

QY 1968 TTNATTACGNATACCATTAAAAAACCCGTATATTTCATATGCGAAAAA 2017

Db 1805 erAsnMetGlnLysAspLeuGluAsnSerAsnAlaLysLeuGlnGluLys 1821

RESULT 11

S26710
spindle pole body protein NUFI - yeast (*Saccharomyces cerevisiae*)
N;Alternate names: Protein D9476.3; protein YDR356w
C;Species: *Saccharomyces cerevisiae*
C;Date: 07-May-1993 #sequence revision 07-May-1993 #text_change 23-Mar-2001
R;Accession: S26710; S34288; A49455; S61152
R;Mirzayan, C.; Copeland, C.S.; Snyder, M.
J. Cell Biol. 116, 1319-1332, 1992
A;Title: The NUFI gene encodes an essential coiled-coil related protein that is
A;Reference number: S26710; PMID:92176232; PMID:1541631
A;Accession: S26710
A;Molecule type: DNA
A;Cross-references: EMBL:X11582; NID:g4069; PIDD:CAA77668.1; PID:g4070
A;Note: the authors translated the codon GAG for residue 206 as Asp and CTG for
R;Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
submitted to the EMBL Data Library, June 1993
A;Description: A spacer element in the *Saccharomyces cerevisiae* spindle pole body
A;Reference number: S34287
A;Accession: S34288
A;Molecule type: DNA
A;Residues: 1-944 <KIL>
A;Cross-references: EMBL:X73297; NID:g312173; PID:g312175
R;Kilmartin, J.V.; Dyos, S.L.; Kershaw, D.; Finch, J.T.
J. Cell Biol. 123, 1175-1184, 1993
A;Title: A spacer protein in the *Saccharomyces cerevisiae* spindle pole body who
A;Reference number: A49455
A;Accession: A49455
A;Molecule type: DNA
A;Residues: 1-22 <KIT>
A;Cross-references: EMBL:X73297
R;Du, Z.
submitted to the EMBL Data Library, June 1995
A;Description: The sequence of *S. cerevisiae* cosmid 9476.
A;Reference number: S61148
A;Accession: S61152
A;Molecule type: DNA

A:Residues: 1-944 <DUZ>
A:Cross-references: EMBL:U28372; NID:G849170; PID:G849173; MIPS:YDR356w
C:Genetics:
A:Gene: SGD:NUF1; SPC110
A:Cross-references: SGD:S0002764; MIPS:YDR356w
A:Map position: 4R
C:Keywords: coiled coil; microtubule; nucleus

Alignment Scores:
Pred. No.: 9.78e-09 Length: 944
Score: 266.00 Matches: 152
Percent Similarity: 41.73% Conservative: 138
Best Local Similarity: 21.87% Mismatches: 225
Query Match: 7.82% Indels: 180
DB: 2 Gaps: 28

US-09-602-362E-15 (1-2030) x S26710 (1-944)

```

QY 150 AAGTCTGTTCCAAATTAAGCCCTTGAATTGAAGAATGAACAAACATTTGAGACGATGAG 209
D 150 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 244 LysThrValLysAspGlnValLeuGluLeuGluAsnAsnSerAspValGlnSer----- 261
QY 210 ATATCCCATCAGATCCCAACAAAGAGACTATGAAGAAAGTCTTGGATCTCAGAGT 269
D 210 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 262 ---LeuLysLeuArgSerLysGluAspGluLeuLysAsnLeuMetAsnGluLeuAsnGlu 280
QY 270 CTCTGTGAGACTGTTTCACAGAGGATGTGTGTTA-----CCCAAGGCTACACATCAA 323
D 270 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 281 LeuLysSerAsnAlaGluGluLysAspThrGlnLeuGluPheLysLysAsnGluLeuArg 300
QY 324 AAAGAAATAGATAAATAATCGAAATTAAGAGAGTCTCTGTAATGATGTTTCTG 383
D 324 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 301 LysArgThrAsnGluLeuAsnGluLeuLysLysLysSerAspGluMetAspLeuGlnLeu 320
QY 384 AAGGCTCCCTGCAGATGAAGTCTTATTCACACTAAAGCCTTA-----GAATGTATG 437
D 384 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 321 Lys-----GlnLysGlnAsnGluSerLysArgLeuLysAspGluLeuAsn 335
QY 438 GACATGCAAACTTTCAAGCAGAGAGCTCCCGAAGAGCCATTCCTTCGAGCTGCCATT 497
D 438 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 336 GluLeuGluThrLysPheSerGluAsnGlySerGlnSerSerAlaLysGluAsnGluLeu 355
QY 498 GAATGCCAAAGCTGTCTCCAAATTAAGCCTTGGATTGAAGATGAACAACATGTGAGA 557
D 498 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 356 LysMetLeuLys-----AsnLysLysLeuGluLeuGluGluGluLysSerThrLys 372
QY 558 GCAGATCATGATTCCTTCACAAATCAAAACAAAGAGGTTGAAGAAATTCCTGGCAT 617
D 558 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 373 AsnSerGlnLeuLeuAlaLysGluGlyLysLeuAlaSerLeuMetAlaGlnLeuThrGln 392
QY 618 TCTGAGAGTCTCGTACAGACTGTTTCACAGAGGATGTGTGTACCCAGGCTACACAT 677
D 618 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 393 LeuGluSer-----LysLeuAsnGlnArgAspSerGlnLeu-----GlySerArg 407
QY 678 CAAAGAAATGATGATAAATAAGTGGAAATTAAGAGATTCAACTAGCTATCAAAATC 737
D 678 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 408 GluGluGluLeuLysThrAsnAspLysLeuGlnLysAspLysLeuAlaArg----- 426
QY 738 TTGGATACAGTTCACTTCTGTGAAGACGAGGAACTTCAAAAAGATCACTGTGAACAA 797
D 738 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 427 GluGluThrValSerLysAspGluArgLysLeuLysLeuLys----- 441
QY 798 CGTACAGGAAATGGAACAATGAAGAGGTTTGTGTACTGAAAGAAAGAACTGTCA 857
D 798 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 442 -----LysValLysGlnLeuGluAsnAspLeuPheValLysLysThrHisSer 458
QY 858 GAAGCAAAA-----GAAATAAAATCACAG-----TTA 884
D 858 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 459 GluSerLysThrIleThrAspAsnGluLeuGluSerLysAspLysLeuLysLysLysLeu 478
QY 885 GAGAACCAA---AAGTT-----AAATGGACAAAGAGCTCTGCGAGT 923
D 885 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 479 GluAsnAspLeuLysValAlaGlnGluLysTyrSerLysMetGluLysGluLeu----- 496

```

```

QY 924 GTGAGATTGACTTTAAACCAGAAAGAGAGAGAAATGCCGATATATTAATGAA 983
D 924 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 497 -----LysGluArgGluPheAsnTyrLysLysLysLysLysLysLysLysLys 508
QY 984 AAAATTAGGAAGAAATTAGGAAGAAATCGAAGAGCAGCATAGGAAGAGATTAGAAGTCAA 1043
D 984 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 509 LysLeuGluAspGluLysThrThrLeuAsnGluLysLysLysLysLysLysLysLysLys 528
QY 1044 CAACAATTGAACAGGCTCTCAGATACAGATACAGATACAGATACAGATACAGATACAG 1103
D 1044 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 529 SerGlnLeuLysAsn-----LysLysLysLysLysLysLysLysLysLysLysLys 537
QY 1104 TTGAATCAGGTTTCTCAGACTCATGAAATGAAATTTATCTCTTACATGAAATTTGCATG 1163
D 1104 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 538 ---AsnSerThrAlaThrHisMetLysLysLysLysLysLysLysLysLysLysLys 549
QY 1164 TTGAAAAGAAATTTGCCATGCTAAAACTGGAATACCCACACTGAAACCAATACACAG 1223
D 1164 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 550 ---GluLysGlnLeuGluSerLeuArgLysAspLysAspLysLysLysLysLysLysLys 568
QY 1224 GAAAGGAAATTAATTAATTTGAG----- 1247
D 1224 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 569 AspSerGluAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 588
D 1248 -----GACATTAGATTTTAAAGAAAGAAATGCTGAATCTCAGATG 1289
D 589 SerGluLysArgSerLysAspLysLysLysLysLysLysLysLysLysLysLysLysLys 608
QY 1290 ACCATAAATGAAGAGGAAATCAATTAATAAGGCGCATCT----- 1331
D 1290 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 609 AsnLeuLysLeuGlnGluAspGluLysSerLysLysLysLysLysLysLysLysLysLys 628
QY 1332 -----CAATATAGTGGCAGCTTAAA 1352
D 1332 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 629 LysAspPheAsnGlnLeuLysSerGluGlnSerAsnLysLysLysLysLysLysLysLys 648
QY 1353 GTTCTGATGCTGAGACACAACTGCTCACTTCTAAATTGAAGAAACCAAGCAAGAA 1412
D 1353 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 649 IleLeuAsnLeuGluAsnLysLeuLysLysLysLysLysLysLysLysLysLysLysLys 668
QY 1413 ATACTAGAGCAGAAATGAATCACACCATCCTAGACTGCTCTGTGTACAGACCAT 1472
D 1413 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 669 SerGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 688
QY 1473 GATCAAAATTGTGACATCAAGAAAAGTCAAGAACTGCTTTCCACATTTGACGAGATGCT 1532
D 1473 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 689 AspArgLeuLeuThrGluLysGluSer-----AlaSerAsp--- 700
QY 1533 TGTTTGCAGAAATGAATGTGTGTAGTAGTACCATATATACAAATGAGGTGCT 1592
D 1533 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 700 ----- 700
QY 1593 CCATCAACCACTTTCTGAAGCTCAAAGAAATCCANAAGCTAAAATTAATCTCAATTA 1652
D 1593 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 701 -----LysGluArgGlu---IleSerIleLeu 708
QY 1653 TGCAGGAGATGCTCTAAGAG-----AAATACATTTGG---TTTCAGGACATGCAAAA 1703
D 1653 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 708 euAsnArgLysLeuAspGluMetAspLysGluLysTrpAsnLeuGlnGluSerLysGluL 728
QY 1704 GAGACCAACGTGAACACACGTCCTCAATGAAGGAGCTGAACACATGTNTCAANCGAAC 1763
D 1704 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 728 yTyrLysArgGluLeuGln---LysValIleThrAlaAsnAspArgLeuArgGluL 747
QY 1764 AAGATNATGTGAACAAACACACATGANCAGCAGGAGTCTCTAGATCAGAAATTTATTTCAA 1822
D 1764 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 747 ySGLuGluLeuAsnGluAsnSerAsnAsnIleArgIleMetGluAspLysMetThrArg 767
QY 1823 -----CTACAAAGCAAAATATGTGCTTCAAC 1850
D 1823 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
D 767 leLysLysAsnTyrLeuSerGluIleThrSerSerLeuGlnGluAsnArgArgLeuGluG 787

```


Db 926 GlnLeuAlaArgSerIleAlaGluGlnThrSerAspLeuGluLysGluLysIleMet 945
Qy 1422 GCAGAAATGAATCAACCATCTAGACTGGCTTCGTGTGAAGACCATCATCAATTT 1481
Db 946 LysGluLeuGlu-----IleLysGluMetMetAlaArgHisLysGlnGlu 960
Qy 1482 GTGACATCAAGAAAAGTCAAGAACCTGCTTTCCACATTCGAGAGATGCTTGTGGAA 1541
Db 961 LeuThrGluLysAspAlaThr-----IleAlaSerLeuGluGluThrAsn 975
Qy 1542 AGAAAAATGAATGTTGATGTGAGTAGTACCGATATATACAAATGAGGTGCTCCATCAACC 1601
Db 976 ArgThrLeuThrSerAspValAlaAsnLeu-AlaAsnGluLysGluGluLeuAsnLys 995
Qy 1602 ACTTCTGAGCTCAAGAAATCCNAAGCTTAAATTAATTCATATTCGAGGAGA 1661
Db 995 sLeuLysGluAlaGlnGluLeuSerArgLeuLys---AspGluGluIleSerAlaAl 1014
Qy 1662 TGCTCTAAGA-----GAAAATACATTTGGTTTCAGGAACATGCACAAAGACCAAC 1712
Db 1014 alaIleLysAlaGlnPheGluLysGlnLeuLeu-----ThrGluArgThrLe 1030
Qy 1713 GTGAACAACAGTGTCAATGAAGAGCTGAACACATGNTCAAACGAAGAAGATNATG 1772
Db 1030 u-LysThrGlnAlaValAsnLysLeuAlaGluIleMetAsnArgLysGluProValLysA 1050
Qy 1773 TGAACAAACACACT-----GANCAGCAGGAGCTCTAGATCAGAAATTA---TTTCAAC 1823
Db 1050 rGlyAsnAspThrAspValArgLysGluLysGluLysGluLysLeuHisMetGluL 1070
Qy 1824 TACAAGCAAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCACATAANGAAAGCTGAC 1883
Db 1070 euLysSerGluArgGluLysLeuThrGlnMetIleLys-----T 1084
Qy 1884 AACAAAGCAAGATACATTCATNTTCATTTCTTGAGAGAAATGCNCATCATCTT- 1942
Db 1084 yrGlnLysGluLeuAsnGluMetGlnAlaGlnIleAlaGluGluSerGlnIleArgIleG 1104
Qy 1943 -----CTAAAGAGAGAAAATGAGAGATATTTNATTACNATAACCATTTAA 1988
Db 1104 LuLeuGlnMetThrLeuAspSerLysAspSerAspIleGluGlnLeuArgSerGlnLeuG 1124
Qy 1989 AAAACCCGTATATTCATA 2008
Db 1124 InAlaLeuHisIleGlyLeu 1130
RESULT 13
A41604
Myosin heavy chain, smooth muscle, long splice form - rabbit
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A41604; A33501
R:Babji, P.; Kelly, C.; Periasamy, M.
Proc. Natl. Acad. Sci. U.S.A. 88, 10676-10680, 1991
A:Title: Characterization of a mammalian smooth muscle myosin heavy chain gene: complete
A:Reference number: A41604; PMID:92073350; PMID:1961735
A:Accession: A41604
A:Molecule type: mRNA
A:Residues: 1-1972 <BAB>
A:Cross-references: GB:M77812
R:Nagai, R.; Kuro-O, M.; Babji, P.; Periasamy, M.
J. Biol. Chem. 264, 9734-9737, 1989
A:Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN
A:Reference number: A33501; PMID:89255535; PMID:2722872
A:Accession: A33501
A:Molecule type: mRNA
A:Residues: 1455-1972 <NAG>
A:Cross-references: GB:J04833; NID:g165519; PIDN:AAA31407.1; PID:g165520
A:Experimental source: smooth muscle
A>Note: examination by Southern blotting for the regions of difference between this isofo
active splicing
C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
F:188-771/Domain: myosin motor domain homology <XMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:559-572/Region: actin binding #status predicted
F:633-647/Region: actin binding #status predicted
F:844-1938/Domain: coiled coil #status predicted <COI>
F:844-1284/Region: S2
F:1285-1972/Region: light meromyosin
F:1939-1972/Domain: carboxyl-terminal <CBT>
F:123/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:701,711/Active site: Cys #status predicted
Alignment Scores:
Pred. No.: 1.86e-08 Length: 1972
Score: 261.50 Matches: 128
Percent Similarity: 40.00% Conservative: 118
Best Local Similarity: 20.81% Mismatches: 215
Query Match: 7.68% Indels: 154
DB: 1 Gaps: 18

US-09-602-362E-15 (1-2030) x A41604 (1-1972)

Qy 528 TTGGAATTTGAAGAAATGAACAA----- 548
Db 827 LeuLysLeuArgAsnTrpGlnTrpTrpArgLeuPheThrLysValLysProLeuLeuGln 846
Qy 549 ---ACATTGAGACAGATCAGATGTTCCCTTCAGATCAAAACAAAGAGGTTGAAGAA 605
Db 847 ValThrArgGlnGluGluMetGlnAlaLysGluAspGluLeuGlnLysIleLysGlu 866
Qy 606 AATTCTTGGGATTCGAGAGTCTCCGTGAGACTGTTTCACAGAGGATGTGTGTGATCCC 665
Db 867 ArgGlnGlnLysAlaGluSerGluLeuGlnGlnLysHisThrGlnLeu--- 885
Qy 666 AAGCTACACATCAAAAGAAATGGATAAAATAGTGGAAATTTAGAAGATTCAACTAGC 725
Db 886 -----SerGluGluLysAsnLeuLeuGlnGlnLeuGlnAlaGluThrGlu 901
Qy 726 CTA-----TCAAAAATCTTG 740
Db 902 LeuTrpAlaGluAlaGluGluMetArgValArgLeuAlaAlaLysLysGlnGluLeuGlu 921
Qy 741 GATACAGTTTCATTCTTGTGAAGAGCAAGGAACTTCAAAAGAGTCACTGTGAAACACGT 800
Db 922 GluIleLeuHisGluMetGluAlaArgLeuGluGluGluAspArgGlyGlnGlnLeu 941
Qy 801 ACAGGAAAAATGGAACAAATGAAAAAGAGAGTGTGTGTACTGAAAGAACTGTGAGAA 860
Db 942 GlnAlaGluArgLysLysMetAlaGlnGlnMetLeuAspLeuGluGlnLeuGluGlu 961
Qy 861 GCAAAAGAAATAAAATCACAGTTTA----- 884
Db 962 GluAlaAlaArgGlnLysLeuGlnLeuGluLysValThrAlaGluAlaLysIleLys 981
Qy 885 -----GAGAACCAAAAGATTAAATCGGAACAA 911
Db 982 LysLeuGluAspAspIleLeuValMetAspGlnAsnAsnLysLeuSerLysGluArg 1001
Qy 912 GAGCTCTGCAGTGTGAGATTGACT-----TTAAACCAAGAGAGAGAGAG 956
Db 1002 LysLeuLeuGluGluArgIleSerAspLeuThrAsnLeuAlaGluGluGluLys 1021
Qy 957 AGAAGAAATCCGATATTTAAATAGAAAATTTAGGAGAGAAATTTAGGAAGATTCGAAGAG 1016
Db 1022 AlaLysAsnLeuThrLysLeuLysAsnLysHisGluSerMetIleSerGluLeuGluVal 1041
Qy 1017 CAGCATAGGAAGAGGTTAGAGTGAACACAACTTGAACAGGCTCTCAGATACAGAT 1076
Db 1042 ArgLeuLysLysGluGluLysSerArgGlnGlnLeuGluLysLysArgLysMetAsp 1061
Qy 1077 ATAGAATTGAAGAGTGTAGAAAAGTAATTTGAATCAGGTTTCTCAGCTCATGAAATGAA 1136
|||

```

Db      1062 GlyAlaSerAsp----- 1066
QY      1137 AATTATCTTACATGAAATTCATGTTGTAAGAAAGAAATTCATGCTAAAGCTGAA 1196
Db      1067 -----LeuHisGlu----- 1077
QY      1197 ATAGCCACACTGAAACACCAATACCAAGAAAGAAATTAATATCTTTGAGGACATTAAG 1256
Db      1078 IleAlaGluLysMetGlnLeuAlaLysGluGluGluAlaLeuAla 1097
QY      1257 AITTTAAAGAAAGATGCTGAATCTCAGATCACCCTAAAG----- 1298
Db      1098 ArgLeuGluAspGluThrSerGlnLysAsnAlaLeuLysLysIleArgGluLeuGlu 1117
QY      1299 -----CTGAAGAGGATCAATTAACATAAGGGCATCTCAATATATAGTGGG 1343
Db      1118 GlyHisIleSerAspLeuGlnGluAspLeuSerGluArgAlaAlaArgAsnLysAla 1137
QY      1344 -----CAGCTTAAGTCTGATAGCTGAG----- 1367
Db      1138 GluLysGlnLysArgAspLeuGlyGluGluLeuGluAlaLeuLysThrGluLeuGluAsp 1157
QY      1368 -----AACCAATGCTCACTCTTAATTAAGGAAAGAAACAAAGAAAGATA--- 1415
Db      1158 ThrLeuAspThrThrAlaThrGlnGlnGluLeuArgAlaLysArgGluGlnGluValThr 1177
QY      1416 -----CTAGAGGAGAAATGTAATCACACCATCTAGACTGCTTCTGCT 1460
Db      1178 ValLeuLysLysAlaLeuAspGluGluThrArgSerHisGluAlaGlnValGlnGluMet 1197
QY      1461 GTACAGACCATGATCAATATGACATCAAGAAAGAAAGTCAAGAAAGCTGCTTCCACATT 1520
Db      1198 ArgGlnLysHisThrGlnValValGluGluLeuThrGluGlnGluGlnPheLysArg 1217
QY      1521 GCAGGAGATGCTTGTTCAGAAAGAAATCAATGTGATGTGAGTAGTACCATATATAA 1580
Db      1218 AlaLysAlaAsnLeuAspLysThrLysGlnThrLeuGluLysGluAsnAlaAspLeu-Al 1237
QY      1581 CAATGAG-----GTGCTCATCAACCATCTTCTGAGGCTCAAGGAATCCANAAGCT 1634
Db      1237 aGlyGluLeuArgValLeuGlyGlnAlaLysGlnGluValGluHisLysLysLysLys 1257
QY      1635 AAAAAATTAATCTC-----AATTATGAGGAGATGCTCTAAGAGAAATATACATT 1682
Db      1257 uGluValGlnLeuGlnGluLeuGlnSerLysCysSerAspGlyGluArgAlaAlaGlu 1277
QY      1683 GGTTCAGGACATGTCACAAAGAGACCACTGAAACACACATGT-----CAAATCA 1733
Db      1277 uLeuAsn-AspLysValHisLysLeuGlnAsnGluValGluSerValThrGlyMetLeu 1297
QY      1734 AGGAAGCTGAA-----CACATGT 1751
Db      1297 erGluAlaGluGlyLysAlaLysLeuAlaLysGluValAlaSerLeuGlySerGlnL 1317
QY      1752 NTCAANCGAACAAAGATNATGTGAACAAACACACTGACGAGGAGTCTTAGATCAGA 1811
Db      1317 euGlnAspThrGlnGluLeuGlnGluGluThrArgGlnLysLeuAsnValSerThrL 1337
QY      1812 AATTATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTCATGCACATA 1871
Db      1337 ysLeuArgGlnLeuGluAspGluArgAsnSerLeuGlnGluGlnLeu-----A 1353
QY      1872 ANGAAGCTGACAAACAAAGCAAGATATACAATGTATNTTCTTCTGAGAGGAAATG 1931
Db      1353 spGluGluMetGluAlaLysGln--Asn-LeuGluArgHisIleSerThrLeuAsnIle 1371
QY      1932 CNCATCATCTTCTAAAGGAAAGATGAGGATATTTTATACNATACCATCAATTTAAAAA 1991
Db      1372 GlnLeuSerAspSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1387
QY      1992 ACCCGTATATTTCAATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2030
Db      1388 ---GluSerLeuGluGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 1399

```

RESULT 14

```

B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Sequence: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: B43402; A43402
R:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
        yosin.
A:Reference number: A43402; MUID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA
A:Residues: 1-2007 <TRAK>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212452
A:Note: The sequence of residues 212-221 and 632-652 and the corresponding nucleotide seq
        uence.
A:Accession: A43402
A:Molecule type: mRNA
A:Residues: 1-211;222-631;653-2007 <TA2>
A:Cross-references: GB:M93676; NID:G212448; PIDN:AAA48988.1; PID:G212449
A:Note: Sequence extracted from NCBI backbone (NCBI:1128664)
C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle myos
        in.
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methylated
        P-loop; Product: myosin heavy chain-B, neuronal #status predicted <MYN>
F:1-211;222-631;653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MYS>
F:188-802/Domain: myosin motor domain homology <MMOT>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:212-221/Region: alternatively spliced segment 1 #status experimental
F:559-593/Region: actin binding #status predicted
F:632-652/Region: alternatively spliced segment 2 #status experimental
F:692-714/Region: actin binding #status predicted
F:875-2007/Domain: coiled coil #status predicted <COI>
F:875-1315/Region: S2
F:1316-2007/Region: light meromyosin
F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:732,742/Active site: Cys #status predicted
F:1854/Binding site: phosphate (Thr) (covalent) #status predicted
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

```

Alignment Scores:

```

Pred. No.: 1,99e-08 Length: 2007
Score: 261.00 Matches: 144
Percent Similarity: 38.81% Conservative: 116
Best Local Similarity: 21.49% Mismatches: 238
Query Match: 7.67% Indels: 172
DB: Gaps: 18

```

US-09-602-362E-15 (1-2030) x B43402 (1-2007)

```

QY      378 TTTCTGAAGCTCCCTCGACAATGAAAGTTTCTTCTCAACTAAAGCTTAGAATGATG 437
Db      369 PheThrLysValLysProLeuLeuGlnValThrArgGlnGluGluGlnAlaLys 888
QY      438 GACATGCAAACTTTCAAAGCAGACGCTCCCGAAGCATCTCGCTTCGAGCCTGCATT 497
Db      889 AspGluGluLeuMetLysValLys-----GluLysGlnThrLysValGluAlaGluLeu 906
QY      498 GAATGCAAAAGTCTGTTCCAATAAAGCTTCGAATTGAAGAT-----GAACAA 548
Db      907 GluGluMetGluArgLysHisGlnGlnLeuLeuGluGluLysAsnIleLeuAlaGluGln 926
QY      549 ACATTGAGACGACATCATGTTTCCCTTCA-----GAATCA 584
Db      927 LeuGlnAlaGluThrGluLeuPheAlaGluAlaGluGluMetArgAlaArgLeuAlaAla 946
QY      585 AAACAAAGAGGTTGAAGAAATTTCTTGGATTCTTGAGAGTCTCCGTGAGACTGTTCA 644
Db      947 LysLysGlnGluLeuGluGluGluLeuLeuHisAspLeuGluSerArgValGluGluGlu 966

```


QY 417 ACTAAAGCCTTAGAATGTGACATGCAAACTTTCAAAGCAGACGCTCCCGAGAGCCA 476
Db 839 ThrLysValProLeuLeuGlnValThrArgGlnGluGluMetGlnAlaLysGlu 858
QY 477 TCTGCCTTCGACCTGCCATTGAAATGCAAAAGTCTGTTCCAAATAAGACCTTGGAATTG 536
Db 859 GluGluMetGlnLysIleThrGluArgGlnGlnLysAlaGluThrGluLeuLysGluLeu 878
QY 537 AAGAATGAACAACAATGTAGACAGCATGATGTTCCCTTCAGATCAAAACAAAGAG 596
Db 879 GluGlnLysHisThrGlnLeuAlaLysLysGlnGluLysThrLeuLeuGlnGluGlnLeuAla 898
QY 597 GTTGAAGAAATCTTGGGATCTGAGAGTCTCCGTGAGACTGTTTCAGAGAGGATGTG 656
Db 899 GluThrGluLeuTyraAlaGluSerGluGluMetArg----- 910
QY 657 TGTGTACCAAGGCTACACATCAAAAGAAATGGATAAAATAGTGGAAATTTAGAAGAT 716
Db 911 ---ValArgLeuAlaLysLysGlnGluLeuGluGluLe----- 923
QY 717 TCAACTAGCCTATCAAAATCTTTGGATACAGTTCATTTCTGTGAAGACGAGGAACCTT 776
Db 924 -----LeuHisGluMetGluAlaArgLeuGluGlu 933
QY 777 CAAAAGATCACTGTGAACAACGCTACAGGAAATGGAACAAATGAAAGAAATTTTGT 836
Db 934 GluGluAspArgGlnGlnLeuGlnAlaGluArgLysLysMetAlaGlnGlnMetLeu 953
QY 837 GTACTGAAAAGAACTGTGAGAAGCAAAAGAAATAAATCACTTA----- 884
Db 954 AspLeuGluGlnLeuGluGluGluGluAlaAlaArgGlnLysLeuGlnLeuGluLys 973
QY 885 -----GAG 887
Db 974 ValThrAlaGluAlaLysIleLysLysLeuGluAspAspIleLeuValMetAspGln 993
QY 888 AACCAAAAAGTTAAATGGGAACAAGAGCTCTGCAGTGTGAGATTGACT----- 935
Db 994 AsnSerLysLeuSerLysGluArgLysLeuLeuGluGluArgValSerAspLeuThrThr 1013
QY 936 ---TTAAACCAAGAACAAGAGAGAGCAAAATGCCGATATATTAATGAATAATTAGG 992
Db 1014 AsnLeuAlaGluGluGluGluLysAlaLysAsnLeuThrLysLeuLysSerLysHisGlu 1033
QY 993 GAAGAATTAGGAAGATCCAGAGACGATAGAAAGAGCTTAGAAGTGAACAACAACACTT 1052
Db 1034 SerMetIleSerGluLeuGluValArgLeuLysLysGluLysSerArgGlnGluLeu 1053
QY 1053 GAACAGGCTCTCAGAATACAGATATAGAAATGAGAGGTGAGAAATTTGATGATCAG 1112
Db 1054 GluLysLeuLysArgLysLeuGlu-----GlyAsp 1063
QY 1113 GTTTCACACTCATGAAATGAAATTTATCTTTACATGAAATTTGATGAAAG 1172
Db 1064 AlaSerAspPheHisGlu----- 1069
QY 1173 GAAATTCCTGCTAAATGGAATAGCACTGCAATCAACACATACAGGAAAGGAA 1232
Db 1070 GlnIleAlaAspLeuGlnAlaGlnIleAlaGluLeuLysMetGlnLeuAlaLysLysGlu 1089
QY 1233 AATAAATCTTTCAGGACATTAAGATTTTAAAGAAAGAAAGTCTGGAATTCAGATGACC 1292
Db 1090 GluGluLeuGlnAlaAlaLeuAlaArgLeuAspGluGluIleAlaGlnLysAsnAla 1109
QY 1293 CTAAAA-----CTGAAAGAGGAATCATTAAC 1319
Db 1110 LeuLysLysIleArgGluLeuGluGluHisIleSerAspLeuGlnGluAspLeuAspSer 1129
QY 1320 AAAGGGCATCTCAATATAGTGG-----CAGCTTAA 1352
Db 1130 GluArgAlaAlaArgAsnLysAlaGluLysGlnLysArgAspLeuGlyGluGluLeuGlu 1149

QY 1353 GTTCTGATAGCTGAG-----AACACAATGCTCACTTCTTAATTGAAG 1394
Db 1150 AlaLeuLysThrGluLeuGluAspThrLeuAspSerIleAlaThrGlnGlnGluLeuArg 1169
QY 1395 GAAAAACAACACAAAGAAATA-----CTAGAGCAGAAATTTGAATCA 1436
Db 1170 AlaLysArgGluGlnGluValThrValLeuLysAlaLeuAspGluGluThrArgSer 1189
QY 1437 CACATCTTAGACTGCTCTGCTGTACAGACCATGTCATTAATTTGACATCAAGAAAA 1496
Db 1190 HisGluAlaGlnValGlnGluMetArgGlnLysHisThrGlnAlaValGluGluLeuThr 1209
QY 1497 AGTCAAGAACTCTTTCCACATTCGAGGAGATGCTTTGTTGCAAGAAAAATGAATGT 1556
Db 1210 GluGlnLeuGluGlnPheLysArgAlaLysAlaAsnLeuAspLysSerLysGlnThrLeu 1229
QY 1557 GATGTAGTAGTACCGATATATAACAATGAG-----GTGCTCCATCAACCTTTCTCA 1610
Db 1230 GluLysGluAsnAlaAspLeu-AlaGlyGluLeuArgValLeuGlyGlnAlaLysGlnG 1249
QY 1611 AGCTCAAGGAAATCCANAAGCCTAAATAATTAATCTC-----AATTATGCAGG 1658
Db 1249 uValGluHisLysLysLysLysLeuGluValGlnLeuGlnAspLeuGlnSerLysCysSe 1269
QY 1659 AGATGCTCTAAGAGAAAAATACATTTGTTTCAGGAACATGCCAAAGAGACCAACGTGAA 1718
Db 1269 rAspGlyGluArgAlaArgAlaGluLeuSer-AspLysValHisLysLeuGlnAsnGluV 1289
QY 1719 CACAGTGT-----CAAATGAAGAAAGCTGAA----- 1744
Db 1289 alGluSerValThrGlyMetLeuAsnGluAlaGluGlyLysAlaIleLysLeuAlaLysA 1309
QY 1745 -----CACATGCTNTCAAANCGAACCAAGATNATGTGAACAACACACTG 1787
Db 1309 spValAlaSerLeuGlySerGlnLeuGlnAspThrGlnGluLeuLeuGlnGluThrA 1329
QY 1788 ANCAGAGAGTCTCTAGATCAGAAATTTTCAACTACAGACCAAAATATGTGCTTC 1847
Db 1329 rgGlnLysLeuAsnValSerThrLysLeuArgGlnLeuGluAspGluArgAsnSerLeuG 1349
QY 1848 AACAGCAATTAGTTTCATGACATAANGAAAGCTGCACAACAAAAAGCAAGATAACATTTGAT 1907
Db 1349 InAspGlnLeu-----AspGluGluMetGluAlaLysGln---Asn-LeuGlu 1363
QY 1908 NTTCAATTCTTCAGAGGAAATGCNCACTCTTCTTAAAGAGAAAAATGAGGAGATAT 1967
Db 1364 ArgHisValSerThrLeuAsnIleGlnLeuSerAspSerLysLysLysLeuGlnAspPhe 1383
QY 1968 TTNATTACNATAACCATTTTAAAAAACCCGTATATTTCAATATATCGAAAAAANAANA 2027
Db 1384 AlaSerThrIle-----GluValMetGluGluGlyLysArgLeuGln 1398
QY 2028 AAA 2030
Db 1399 Lys 1399

Search completed: July 15, 2004, 09:47:57
Job time : 105.671 secs

; PRIOR APPLICATION NUMBER: US 60/294,443
 ; PRIOR FILING DATE: 2001-05-29
 ; NUMBER OF SEQ ID NOS: 141
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1341
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-058-270A-4

Alignment Scores:

Pred. No.: 4,45e-233 Length: 1341
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 12 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-058-270A-4 (1-1341)

QY 9 GTTAAAGATGCTCTTCTGAGGCTAACTCGGAAATGAAAGTTTCTATTCCAACTAAAGCC 68
 Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCCTCCCGAGAAGCCATCTGCCTTC 128
 Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705
 QY 129 GAGCCTGCCATGAAATGCAAAAGTCTGTCMAATAAAGCCCTTGAAGTTGAAGATGAA 188
 Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
 QY 189 CAAACATTGAGACGACATGACATCTCCCATCAGAACTCCAAACAAAGAGACTATCAAGAA 248
 Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
 QY 249 AGTCTTGGGATCTCTGAGAGTCTCTGAGAGCTGTTTCAGAAGAGTGTGTGTTACCC 308
 Db 746 SerSerTrpAspSerGluSerLeuLysGluThrValSerGlnLysAspValCysLeuPro 765
 QY 309 ARGCGTACACATCAAAAGAAATAGATAAAATAATGGAATAATGAGAGAGTCTCCTGAT 368
 Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLysProAsp 785
 QY 369 AATGATGGTTTCTGAAGCTCCCTCCAGATCAAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
 QY 429 GAATTCATGACATGCAAACTTTCAAGCAGAGCCCTCCCGAGAAGCCATCTGCCTTCGAG 488
 Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 825
 QY 489 CCTGCCATTGAATGCAAAAGTCTGTTCCAAATAAAGCCCTTGAATTTGAAGAATGAACAA 548
 Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
 QY 549 ACATTGAGACGACATGAGATGTTCCCTTCAGATCAAAACAAAGAGTTTGAAGAAAT 608
 Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
 QY 609 TCTTGGGATCTCAGAGTCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTACCCCAAG 668
 Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
 QY 669 GCTACACATCAAAAGAAATGGATAAATAAGTGGAAATTTAGAAGATTCAACTAGCCTA 728
 Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
 QY 729 TCAAAATCTTGATACAGTTCTTCTGTGAAAGCAAGGAACTTCAAAAGATCAC 788
 Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
 QY 789 TGTGAACACGTTACAGAAATAATGGAAACAAATGAAAGAGTTTTGTGTACTGAAAGAG 848

Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
 QY 849 AAATGTCAGAGCAAAAGAAATAAATACAGTTAGAGAACCAAAAGTTAAATCGGAA 908
 Db 946 LysLeuSerGluAlaLysGluIleLysSerGluLeuGluAsnGlnLysValLysTrpGlu 965
 QY 909 CAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAAGAGAGAGAAAGAAATGCC 968
 Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
 QY 969 GATATATTAATAACAAAATTTAGCGAAGAAATTAGGAAGAAATCGAAGCAGCATAGAAA 1028
 Db 986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 1005
 QY 1029 GAGTTAGAAGTGAACCAACACTTGAACAGGCTCTCAGAAATACAAGATATAGAATGAAG 1088
 Db 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
 QY 1089 AGTGTGAAAAAGTAATTTGAATCAGGTTTCTCACACTCATGAAATGAAAATTTATCTCTTA 1148
 Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
 QY 1149 CATGAAAATTTGCATGTTGAAAAAGGAATTCCTCATCTTAAACTGAAATAGCCACATG 1208
 Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
 QY 1209 AATCACCATAATACAGAGAAAGAAATAATATCTTTGAGACATTAAGATTTTAAAGAA 1268
 Db 1066 LysHisGlnTyrGlnGlnLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
 QY 1269 AAGAACTCTGAATCTCAGATGACCTTAAACTGAAAGAGAAATCATTAATAAGCGCA 1328
 Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1105
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGACCAATGCTCACTTCTAAA 1388
 Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
 QY 1389 TTGAAGAAAAACAAGCAAAAGAAATCTAGAGCGCAGAAATGAATCAGACCATCTCTAGA 1448
 Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 1145
 QY 1449 CTGCTTCTCTCTGACAGACCATGATCAATTTGTCACATCAAGAAAAAGTCAAGAACCT 1508
 Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
 QY 1509 GCTTTCACACTTCGAGGAGATGCTTGTTCCTCAAGAAAAATGAATGTTGATGTGAGTAGT 1568
 Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
 QY 1569 ACCGATATATAACATGAGTCTCCATCAACACTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
 QY 1629 AAGCTAAAAATTAATCTCAATTTATGAGGAGATGCTTCAAGAGAAAAATACATGTTGTTTC 1688
 Db 1205 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1225
 QY 1689 AGGAACATGCAAAAGAGACCAACGTTGAACACACAGTGTCAATGAAGAGAGCTGAACACA 1748
 Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHis 1245
 QY 1749 TGTNTCAAAACGCAAGATNATGTGAACAAACACACTGCANCAAGCAGGAGTCTCTAGATC 1808
 Db 1245 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
 QY 1809 AGAAATTTTCAACTACAAAGCAAAATAATGTGGCTTCAACAGCAATTAAGTTCATGCAC 1868
 Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHis 1285
 QY 1869 ATAANGAAAGCTGACCAACAAAGCAAGATAAACAATGATNTTTCATNTTCTTGAGAGAAA 1928

```
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTTCTAAAGAGAAATGAGGAGATATTNATACNATAACCATTTA 1987
Db 1305 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAAACCCGCTATATTCAATATGGAAGAAAAAANAANA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337

RESULT 2
US-10-007-805-565
; Sequence 565, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-007-805-565

Alignment Scores:
Pred. No.: 4,456-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.4% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-565 (1-1341)
QY 9 GTTAAAGATGPTCTTCTGAAGGCTAACTGCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTAGAATTGATGCATGCAAACTTTCAAAGCAGAGCTCCCGAAGACCACTATGCGCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 705
QY 129 GAGCTGCAATGAATGCAAGCTGTTCCAAATGAAGCTTGGATTGAGATGAA 188
Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAACATTCAGACGACATGAGATATCCCATCAGAAATCAACAAAGGACTATGAAGAA 248
Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY 249 AGTTCTGGATTCTGAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTACCC 308
Db 746 SerSerTrpAspSerGluSerLeuLysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATGAAATGAAATGAAAGAGTCTCCTCAT 368
Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
```

```
QY 369 AATGATGTTTCTGAAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGACCACTATGCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTGCGCAATTGAAATGCAAAAGTCTGTCCAAATTAAGACCTTGGAATTTGAAGAATGAACA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGAAATCAAAAAGGAGGTTGAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 509 TCTGGGATTTCTGAGATCTCCGTGAGACTGTTTTCAGAGAGGATGTGTGTACCCAAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAAGAAATGGAATAAATGGAATAAATGGAATAAATGGAATAAATGGA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGATACAGTTTCATTTCTGTAAGAGCAAGGAACTTCAAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACAACGTACAGAAAAATGGAACAAATGGAACAAATGGAACAAATGGAACAAATGGA 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAAGTGTGAGAACAAAGAAATAAATACAGTTAGAGAACCAAAAGTAAATGGAAGAA 908
Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAGAGAGATGCC 968
Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
QY 969 GATATATTAATGAAATAATTAGGGAAGAAATAGGAAGAAATAGGAAGAGCAGCAGTAA 1028
Db 986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 1005
QY 1029 GAGTTAGAAGTGAACAACTTGAACAGGCTCTCAGATACAGATACAGATACAGATACAG 1088
Db 1006 GluLeuGluValLysGlnGlnLeuGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
QY 1089 AGTGTAGAAAGTAAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATTTATCTCTTA 1148
Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
QY 1149 CATGAAATTCATGTTGAAAAAGGAAATTCGCATGCTTAAATCGAAATAGCACACTG 1208
Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACAGAGAAAGGAAATAAATCTTTGAGCAGATTAAGATTTTAAAGAA 1268
Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLeuLysGlu 1085
QY 1269 AAGAATGCTCAACTCAGATGACCCCTAAATCTGAAAGAGAAATTCATTAATAAAGGCA 1328
Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACACATGCTCACTTCTAA 1388
Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGAAAAACAAGACAAAGAAATACTAGAGCGCAAAATTCGAATCACACCATCTCTAG 1448
Db 1126 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisSerProArg 1145
QY 1449 CTGCTTCTCTCTGTACAGACCATGATCAAAATTTGTGACATCAAGAAAAAGTCAAGAACCT 1508
```

```
Db      1146 LeuAlaSerAlaValGlnAspHisGlnIleValThrSerArgLysSerGlnGluPro 1165
QY      1509 GCTTTCCACATTCGACGAGATGCTGTTTGCAGAAAGAAAATGAATGTTGATGTAGTAGT 1568
Db      1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY      1569 ACCGATATATAAATAGAGTCTCATCAACACACTTTCTGAAGCTCAAGGAAATCCAN 1628
Db      1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
QY      1629 AAGCCTAAATAATTAATCTCAATATATCCAGGAGATGCTCTAAGAGAGAAAATACATTTGTTTC 1688
Db      1205 SSeLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 1225
QY      1689 AGGAACATGCACAAAGAGACCAACGTGAACACAGTGTCAATAGAGAGAGCTGAACACA 1748
Db      1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
QY      1749 TGNNTCAAAACGAAACAGATNATGTGAACAAACACACACTGANCAGCAGGAGTCTCTAGATC 1808
Db      1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY      1809 AGAAATTAATTCACACTACAAACAAATATATGTGCTTCAACAGCAATTAGTTCATGCAC 1868
Db      1265 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
QY      1869 ATAAGAAACCTGACACAAACAAACAGATACAAATGATNTCAATNTCTTCTGAGAGGAAA 1928
Db      1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY      1929 ATCCN-CATCATCTTCTAAAGAGAGAAAATGAGGAGATATTTNATTACNATAACCATTTA 1987
Db      1305 MetGlnHisIleLeuLysGluLysAsnGluGluLysPheAsnTyrAsnAsnHisLeu 1324
QY      1988 AAAAAACCGTATTTCAATATGGAAGAAAAAANAANAANA 2027
Db      1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337

RESULT 3
US-10-076-622-565
; Sequence 585, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRI
; ORGANISM: Homo sapiens
US-10-076-622-565

Alignment Scores:
Pred. No.: 4,45e-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 14 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-076-622-565 (1-1341)
QY      9 GTTAAAGATGGTCTTCTGAGGCTACTCGGAATGAAATGTAATCTTATCCCACTAAAGCC 68
Db      666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
```

```
QY      69 TTAGAATTGATGACATGCAAACTTTTCAAGCAGAGCCCTCCGAGAGCAATCTGCTTC 128
Db      686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGlnLysProSerAlaPhe 705
QY      129 GAGCCTCCCAATTCGAATGCAAAAGTCTGTTTCCAAATAAAGCCCTTGGAAATGAAGATGAA 188
Db      706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY      189 CAACATTGACAGACAGATGATCTCCCATCAGATCCAAATCCAAACAAAGAGACTATGAGAA 248
Db      726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY      249 AGTCTTGGAGTCTGAGACTCTCTGTGAGACTGTTTTCACAGAGAGATGTGTGTTTACC 308
Db      746 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
QY      309 AAGCTACACATCAAAAGAAATAGATATAAATAATGGAATAATAGAGAGTCTCTGAT 368
Db      766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 785
QY      369 AATGATGCTTTCTGAGGCTCCCTGCAGATGAAAGTCTTATTCACAACTAAAGCCTTA 428
Db      786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY      429 GAATTGATGACATGCAAAACCTTTCAAAGCAGAGCCCTCCGAGAGCCATCTGCTTCGAG 488
Db      806 GluLeuMetAspMetGlnThrPheLysAlaGluProGlnLysProSerAlaPheGlu 825
QY      489 CTTGCCATTCGAATGCAGAGTCTGTTTCCAAATAAAGCCCTTGAATTAAGAGATGAACAA 548
Db      826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysGlnGlu 845
QY      549 ACATTGAGACAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGAGTTCAGAGAAAT 608
Db      846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 865
QY      609 TCTTGGATTCGAGAGTCTCCGTGAGCTGTTTTCAGAGAGAGTCTGCTGTACCCAG 668
Db      866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY      669 GCTACACATCAAAAGAAATGGAATAAATAAGTGAATAATAGAGATTCACATGACCTTA 728
Db      886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY      729 TCAAAATCTTGATACAGTTCATTTCTGTAAGAGAGAGGAACTTCAAAAGATCAC 788
Db      906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGlnLeuGlnLysAspHis 925
QY      789 TGTGAACACGTACAGGAAAAATGGAACAAATCAAAAGAAAGTGTGTGTACTGAAAAAG 848
Db      926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 945
QY      849 AATCTGTCAAGACAAAGAAATATAATACACTTAGAGAACCAAAAGTTAAATGGGAA 908
Db      946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY      909 CAGAGCTCTGCAAGTGTGAGATTGACTTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
Db      966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
QY      969 GATATATTAATGAAAAATATAGCGAAGATTAAGGAAGATTCAGAGCAGCATAGGAAA 1028
Db      986 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 1005
QY      1029 GAGTTAGAGTGAACACAACTTGAACAGGCTCTCAGATACAGATATAGATTAAG 1088
Db      1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1025
QY      1089 AGTGTAGAAAGTAAATTTGAATCAGGTTTCTCACACTCATGAAAAATGAAATTAATCTCTTA 1148
Db      1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
```


QY 429 GAATTGATGACATGCAAACTTCAAGCAGAGCTCCGAGAACCACTGCTCCTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CTGCGCAATGAATGCAAAAGCTGTTCACAAATAAGCCTTGAATGAAGAAATCAACAA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAlaLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTCAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTGGGATCTCAGAGTCCGTGAGACTGTTTCACAGAGGATGTGTGTCACCCAG 668
Db 866 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGGATAAAATAAGTGGAAATTAGAAGATTCAACTAGCCTA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 905
QY 729 TCAAAATCTTGATACAGTTCTTCTGAAAGAGCAAGGAATTCACAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGAACACAGTACAGAAATGGAAACAAATCAAAAGCAAGTTTGTGTACTGAAAG 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAATCTCAGAACCAAGAAATAAATCACAGTTAGAGAACCAAAAGATTAAATGGGAA 908
Db 946 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 965
QY 909 CAGAGCTCTGAGTGTGAGTGTGACTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 985
QY 969 GATATATTAATGAAATAGGAGAAATAGGAGAAATAGGAGAGATAGGAGAA 1028
Db 986 AspIleLeuAsnGlnLysIleArgGluLeuGluArgIleGluGlnHisArgLys 1005
QY 1029 GAGTTAGAGTGAACCAACACTGACAGGCTCTCAGATACAGATACAGATTAAGTGAAG 1088
Db 1006 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLysLys 1025
QY 1089 AGTGTAGAAGTAATTTGAATCAGGTTTCTCACACTCATGAAATGAAATATCTCTTA 1148
Db 1026 SerValCysSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTrpLeuLeu 1045
QY 1149 CATGAAATTCATGTTGAAAGAAATGCAATGCAATGCAATGCAATGCAATGCAATG 1208
Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065
QY 1209 AAACACCAATACCAAGAAAGGAAATTAATCTTTGAGACATTAAGATTTTAAAGAA 1268
Db 1066 LysHisGlnTrpGlnGlnLysGluAsnLysTrpPheGluAspIleLysIleLeuLysGlu 1085
QY 1269 AAGAATCTGAACTCAGATGACCCCTAAACCTGAAAGAGAAATCAATTAACAAAGGCA 1328
Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGGAGCTTAAGTCTGATAGCTGAGACACAAATGCTCACTCTCAA 1388
Db 1106 SerGlnTrpSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGGAAACCAAGCAAGAAATCTAGAGGCAAGAAATTAATCAACCATCTCTAGA 1448
Db 1126 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1145
QY 1449 CTGCTCTCTGCTGACAGACCATGATCAATGTCACATCAAGAAAGTCAAGAACT 1508
Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
QY 1509 GCTTTCCACATTGAGGAGATGCTTTTTCGAAAGAAATGAATGTTGATGTGAGTAGT 1568

Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATAACAATGAGGTGCTCCATCAACACACTTTCTGAAGCTCAAGGAAATCCAN 1628
Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1205
QY 1629 AGCCTTAATAATTAATCAATTCAGAGATGCTCTAAGAGAAATACATTTGGTTTC 1688
Db 1205 sSerLeuLysIleAsnLeuAsnTrpAlaGlyAspAlaLeuArgLysGlnThrLeuValSe 1225
QY 1689 AGGAACATGCACAAAGAGACCAACAGTCAACAGTGTCAAAATGAAGAGAGCTGAACACA 1748
Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
QY 1749 TGNVTCMAANGCAAGATNATGTGAACAAACACTGANCAGCAGGAGTCTCTAGATC 1808
Db 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY 1809 AGAATTTATTTCAACTACAAAGCAAAATATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
Db 1265 lnyLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATTAACAAATTTGATNTTCATTTCTTGAGAGAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTTCTAAAGAGAAATAGAGAGATTTTATTCATACATTAACCACTTTA 1987
Db 1305 MetGlnHisLysLeuLeuLysGluLysAsnGluGluIlePheAsnTrpAsnAsnHisLeu 1324
QY 1988 AAAAACCCGATATTTCAATATGCAATATGCAAAACCAAAACCAAAACCAAAACCA 2027
Db 1325 LysAsn-ArgIleTyrGlnTrpGluLysGluLysAlaGlu 1337
RESULT 5
US-10-124-805-565
; Sequence 565, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: 210121.470C12
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ. ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-124-805-565
Alignment Scores:
Pred. No.: 4,45e-233 Length: 1341
Score: 3272.00 Matches: 656
Percent Similarity: 97.49% Conservative: 3
Best Local Similarity: 97.04% Mismatches: 13
Query Match: 96.15% Indels: 5
DB: 14 Gaps: 0
US-09-602-362E-15 (1-2030) x US-10-124-805-565 (1-1341)
QY 9 GTTAAAGATGCTCTTCTGAAGGCTAACTCGGAAATGAAGCTTCTATTCCAACTAAGCC 68
Db 666 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 685
QY 69 TTCAATTTGATGACATGCAAACTTTTCAAGCAGAGCTCCCGAAGAGCCATCTGCTTC 128
Db 686 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 705

QY 129 GAGCTGCGCATGAAATGCAAAAGTCTGTTCCTCAATATAAGCCCTTGAAATTGAAGATGAA 188
Db 706 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 725
QY 189 CAAACATTCAGAGCAGATGAGATACCTCCATCAGAAATCCAAACAAAGAGACTATGAAGAA 248
Db 726 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 745
QY 249 AGTCTCTGGGATTCTGAGAGTCTCTGTGAGACTGTTCACAGAGAGTGTGTTCACCC 308
Db 746 SerSerTyrPaspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 765
QY 309 AGGCTACACATCAAAAAGAAATAGATAAAATATAATGGAATAATTAGAAGACTCTCTCAT 368
Db 766 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 785
QY 369 AATGATGGTTCCTGAGGCTCCCTGCAGAAATGAAAGTTTCTATTCCCACTAAAGCCTTA 428
Db 786 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 805
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAGAAGCCCATCTGCCTTCGAG 488
Db 806 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 825
QY 489 CCTGCAATGAAATGCAAAAGTCTGTTCCTCAATATAAGCCCTTGAAATGAAGATGAAACA 548
Db 826 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 845
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 846 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysValGluGluAsn 865
QY 609 TCTTGGGATTCAGAGTCTCGTGAGACTGTTTCACAGAGAGTGTGTGTACCCAAAG 668
Db 866 SerTyrPaspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 885
QY 669 GCTACACATCAAAAGAAATGATATAATAGTGAAGAAATAGAGATTCACATAGCCTA 728
Db 886 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluLysSerThrSerLeu 905
QY 729 TCAAAATCTTCGATACAGTTCATCTGTGAAAGAGCAAGGAACTTCAAAAGATCAC 788
Db 906 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 925
QY 789 TGTGACACAGTACAGGAAATGCAACAAATGAAAGAAAGTGTGTCTACTGAAAG 848
Db 926 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 945
QY 849 AAACCTGTGAGAGCAAAAGAAATAGGCAAGAAATAGGCAAGAAATAGGCAAGCAATAGGAA 1028
Db 946 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTyrGlu 965
QY 909 CAAGAGCTCTGAGTGTGAGATGCTTTAACTCAAGAGAGAGAGAGAGAAATGCC 968
Db 966 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgAsnAla 985
QY 969 CATATATTAAATGCAAAATAGGCAAGAAATAGGCAAGAAATAGGCAAGCAATAGGAA 1028
Db 986 AspileLeuAsnGlnLysIleArgGluLeuGluArgIleGluGluGlnHisArgLys 1005
QY 1029 GAGTTAGAAGTGAACAAACACTCAAGAGGCTCTCAGAAATACAGATATAGAAATGAAG 1088
Db 1006 GluLeuGluValLysGlnLeuGluGlnAlaLeuArgIleGlnAspileGluLeuLys 1025
QY 1089 AGTGTAGAAAGTAAATTCGATCAGGTTCTCACACTCATGAAATGAAATATCTCTTA 1148
Db 1026 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1045
QY 1149 CATGAAATTCGATGTTGAAAGGAAATTCGCATGCTTAAAGTGAATAGCCACACTG 1208
Db 1046 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1065

QY 1209 AAACACCAATACCAGGAAAGGAAATAATACTTTGAGGACATTAAGATTTTAAAGAA 1268
Db 1066 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1085
QY 1269 AAGAACTCTGAACCTTCAGATGACCTTAAACTGAAAGAGGAAATCATTAATAAAGCGCA 1328
Db 1086 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLeuSerLeuThrLysArgAla 1105
QY 1329 TCTCAATATAGTGGCAGCTTAAAGTCTCATAGCTGAGAACACAATGCTCCTCTTAA 1388
Db 1106 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1125
QY 1389 TTGAAGAAACAAAGACAAAGAAATACTAGAGCGAGAAATGAATCACCACCATCTCTAGA 1448
Db 1126 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 1145
QY 1449 CTGCTTCTCTCTGACAGAGCCTGATCAATTTGTGCATCAAGAAAAGCTCAAGAACCT 1508
Db 1146 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1165
QY 1509 GCTTTCACATTCGAGGAGATGCTTGTTCGAAAGAAAAATGAATGTGTGAGTAGT 1568
Db 1166 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1185
QY 1569 ACCGATATATAACATGAGGTGCTCCATCAACACACTTCTGAGGCTCAAGAGAAATCCAN 1628
Db 1186 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy 1205
QY 1629 AAGCTAAAAATTAATCTCAATATGAGGAGATGCTCTAAGAGAAAAATACATTGGTTTC 1688
Db 1205 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1225
QY 1689 AGGAACTGCAAAAGAGACCAACCTGAAACACACAGTGTCAAAATGAAGGAAGCTGACACA 1748
Db 1225 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1245
QY 1749 TGTNTCAAAACGACAGATNATGTGCAACAAACACACTGANCAGCAGGAGCTCTAGATC 1808
Db 1245 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1265
QY 1809 AGAAATTTTCAACTACAAAGCAAAATAATGTGGCTTCAACAGCAATTAGTTTCATGCAC 1868
Db 1265 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1285
QY 1869 ATAANGAAGCTCACACAAACAGATACAAATGATTTTTCATTTCTTGAGAGAA 1928
Db 1285 is-LysLysAlaAspAsnLysSerLysIleThrIleAspileHisPheLeuGluArgLys 1304
QY 1929 ATGCN-CATCATCTTCTTAAAGAGAAAAATGAGGAGATATTTNATTACNATAACCATTTA 1987
Db 1305 MetGlnHisHisLeuLeuLysGlnLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1324
QY 1988 AAAACCCGTATATTCATATATGGAAGAAAAAANAANA 2027
Db 1325 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1337

RESULT 6

US-10-007-805-573
; Sequence 573, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
 ; FILE REFERENCE: 210121.470C10
 ; CURRENT APPLICATION NUMBER: US/10/007,805
 ; CURRENT FILING DATE: 2001-12-07
 ; NUMBER OF SEQ ID NOS: 593
 ; SOFTWARE: fastseq for Windows Version 4.0
 ; SEQ ID NO 573
 ; LENGTH: 1349
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-007-805-573

Alignment Scores:
 Pred. No.: 4,45e-233 Length: 1349
 Score: 3272.00 Matches: 656
 Percent Similarity: 97.49% Conservative: 3
 Best Local Similarity: 97.04% Mismatches: 13
 Query Match: 96.15% Indels: 5
 DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-573 (1-1349)

QY 9 GTTAAAGATGCTCTCTGGAAGCTAACTCGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
 Db 674 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 693
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAGCAGAGCCTCCGAGAGCCATCTGCCTTC 128
 Db 694 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 713
 QY 129 GACCTGCCATTCAAAATGCAAAAGTCTGTTCCAAATAAAGCCTTGCAATTGAAGAAATCAA 188
 Db 714 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 733
 QY 189 CAAACATTGAGACGACATGAGATCTCCATCAGAAATCCAAACAAAGACATATGAAGAA 248
 Db 734 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 753
 QY 249 AGTCTTGGGATCTCAGAGTCTCTGAGACTGTTTTCAGAAAGGATGTGTGTTTACCC 308
 Db 754 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 773
 QY 309 AAGCTACACATCAAAAGAAATAGATAAAATAAATGGAATTTAGAGAGTCTCTGTAT 368
 Db 774 LysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLysProAsp 793
 QY 369 AATGATGTTTCTGAGGCTCCCTCAGAAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
 Db 794 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 813
 QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCGAGAGCCATCTGCCTTCGAG 488
 Db 814 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 833
 QY 489 CTTGCCATTCGAATGCAAAAGTCTGTTCCAAATAAAGCCTTGAAATTGAAGAAATGAACA 548
 Db 834 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 853
 QY 549 ACATTGAGACGACATCAGATGTTTCCCTTCAGAAATCAAAACAAAGAGGTTTGAAGAAAT 608
 Db 854 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLysValGluGluAsn 873
 QY 609 TCTTGGGATCTCAGAGTCTCCGTGAGACTGTTTCAGAGAGGATGTGTGTATCCCAAG 668
 Db 874 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 893
 QY 669 GCTACACATCAAAAGAAATGGAATAAATGGAATTTAGAAATTTCAACTAGCCTTA 728
 Db 894 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 913
 QY 729 TCAAAATCTTGATACAGTTCCTTCTGTGAAGAGCAGGGAATCTCAAAAGATCAC 788
 Db 914 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuLysAspHis 933

QY 789 TGTGAACACAGTACAGAAAAAATGGAACAAATGGAAGAAAGTTTGTGTACTGAAAAAG 848
 Db 934 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 953
 QY 849 AAATCTGTCAAGAACAAAAAATAAATCACAGTTAGAGAACCAAAAAAGTTAAATCGGAA 908
 Db 954 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 973
 QY 909 CAAGAGCTCTGAGTGTGAGATTGACTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 Db 974 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 993
 QY 969 GATATATTAATCAAAAAATTAGGGAAGAAATAGGAAGAAATCAAGACAGCATAGAAA 1028
 Db 994 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGluGlnHisArgLys 1013
 QY 1029 GAGTTAGACTGAAACAACTTGAACAGGCTCTCAGAAATACAGATATAGAAATTGAAG 1088
 Db 1014 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 1033
 QY 1089 AGTGTAGAAAGTAATTTGAATCAGGTTCTCACATCATGAAATGAAATATCTCTTTA 1148
 Db 1034 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 1053
 QY 1149 CATGAAATTCATGTTGAAAAAGGAAATTCCTCATCTGCTAAACTGAAATAGCACACTG 1208
 Db 1054 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 1073
 QY 1209 AAACACCAATACAGGAAAGGAAATTAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 1074 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysLeuLysGlu 1093
 QY 1269 AAGAAATCTCAACTTCAGATGACCTTAAACTGAAAGAGGAATCATTAATAAAGGCA 1328
 Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluSerLeuThrLysArgAla 1113
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGACACAACTGCTCCTCTTAA 1388
 Db 1114 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1133
 QY 1389 TTGAAGCAAAAAACAAGCAAAAGAAATACTAGAGGCAAAATTCGAATCAACACCATCTAGA 1448
 Db 1134 LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluLeuGluSerHisProArg 1153
 QY 1449 CTGCTTCTGCTGTACAAGCCATGATCAAAATTTGTGACATCAAGAAAGTCAAGAACCT 1508
 Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
 QY 1509 GCTTTCACATTCAGAGAGATGCTTGTTCGAAAGAAAAATGAATGTGTGAGTAGT 1568
 Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGluArgLysMetAsnValAspValSerSer 1193
 QY 1569 ACCGATATATAACAAATGAGTGTCTCCATCAACCACTTCTGAAGCTCAAGGAAATCCAN 1628
 Db 1194 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
 QY 1629 AAGCTTAAAAATTAATCTCAATATGAGGAGATGCTCTAAGAGAAAAATACATTGGTTTC 1688
 Db 1213 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1233
 QY 1689 AGGAACATGCACAAAGAGACCAACGTCGAAACACACAGTGTCAAAATGAAGAAAGTGAACACA 1748
 Db 1233 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
 QY 1749 TGTNTCAAAACCAACAGATNATGTGAACAAACACTGANCAGCGAGTCTCTAGATC 1808
 Db 1253 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnGlnGlnGln 1273
 QY 1809 AGAAATATTTTCAACTACAAAGCAAAAAATATGTGGCTTCAACAGCAATTAGTTCATGCAC 1868
 Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 1293

Db 1074 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 1093
QY 1269 AAGAACTCTGAATCTAGATGACCTAAACTGAAAGAGAGAAATCACTAAAGGGCA 1328
Db 1094 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 1113
QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACCAATGCTCACTCTAAA 1388
Db 1114 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 1133
QY 1389 TTCAAGGAAAAACAAGCAAGAAATCTAGAGCGAGAAATGAATCACCACTCTAGA 1448
Db 1134 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 1153
QY 1449 CTGGCTTCTGCTCAACAAGACCATGATCAAAATTTGTCACATCAAGAAAAAGTCAAGAACT 1508
Db 1154 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 1173
QY 1509 GCTTTCACATTCAGAGAGATGCTTCTTTGCAAGAAAAATGAATGATGATGATGAT 1568
Db 1174 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 1193
QY 1569 ACCGATATATAACAATGAGTCTCCATCAACCACTTTCTGAAGCTCAAGCAAAATCCAN 1628
Db 1194 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 1213
QY 1629 AAGCCTTAAATAATTAATCTCAATATGTCAGGAGATGCTCTAAGAGAAAAATACATTTGTTTC 1688
Db 1213 SserLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSe 1233
QY 1689 AGGACATGCACAAGAGACCAAGCTGAACACACAGTGTCAATGAAGAGAGCTGAACACA 1748
Db 1233 I-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 1253
QY 1749 TGTNTCAAAACGAAACAAAGATNATGCAACAAACACACTGACAGAGAGTCTAGATC 1808
Db 1253 eTyr-GlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 1273
QY 1809 AGAATATTTCAACTACAAACAAATAATGCTTCAACAGCAATTAGTTCATGCAC 1868
Db 1273 InLysLeuPheGlnLeuGlnSerLysAsnMetTyrLeuGlnGlnGlnLeuValHisAlaH 1293
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATACAAATGATNTTCAATNTCTTGAGAGGAAA 1928
Db 1293 Is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 1312
QY 1929 ATGCN-CATCACTCTTAAAGAGAAAAATGAGAGATATTNATTACNATAACCATTTA 1987
Db 1313 MetGlnHisHisLeuLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 1332
QY 1988 AAARACCCGTATTTCAATATGGAAGAAAAAANAANAANA 2027
Db 1333 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 1345

RESULT 9

US-09-604-287A-475
; Sequence 475, Application US/09604287A
; Patent No. US20020064872A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Dillon, David C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 475

; LENGTH: 1002
; TYPE: PRT
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)-(1002)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-604-287A-475

Alignment Scores:
Pred. No.: 2,78e-232 Length: 1002
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 9 Gaps: 0

US-09-602-362E-15 (1-2030) x US-09-604-287A-475 (1-1002)

QY 9 GTTAAAGATGCTCTCTGAAGGCTAACTGCGGATGAAGTTCTATTCCAACATAAGCC 68
Db 327 VallysaspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
QY 69 TTGAATTTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGACCAATCTGCCTTC 128
Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 366
QY 129 GAGCTCCCATTTGAAATGCAAAAGTCTGTTCCAAATTAAGCCTTGGAATTAAGAAATGAA 188
Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
QY 189 CAAACATTTGAGAGCAGATGAGATCTCCCATCAGATCCCAACAAACAAAGAGACTATGAAGAA 248
Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
QY 249 AGTTCTTGGGATTTCTGAGATCTCTGTGAGACTGTTTCACAGAGAGATGTGTGTTTACCC 308
Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
QY 309 AAGCTACACATCAAAAGAAATAGATAAATAAATGAAATTAAGAAATTAAGAGAGTCTCTGAT 368
Db 427 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluSerProAsp 446
QY 369 AATGATGTTTCTGAAGGCTCCCTCGAGATGAAAGTTTCTATTCCAACTAAAGGCTTA 428
Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGACCAATCTGCCTTCGAG 488
Db 467 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 486
QY 489 CCTGCCATTGAAATGCAAAAGTCTGTTCCAAATTAAGGCTTGAATTAAGAAATGAACAA 548
Db 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGACGACATCAGATGTTCCCTTCAGAAATCAAAACAAAGAGGTTTGAAGAAAT 608
Db 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATTTCTGAGAGTCTCCGAGAGCTGTTTCACAGAGAGATGTGTGTACCCAAG 668
Db 527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
QY 669 GCTACACATCAAAAGAAATGGAATAAATAAGTGGAAAAATTAGAAGATTCAATAGCCTTA 728
Db 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
QY 729 TCAAAATCTTGATACAGTTCATCTTGTGAAGAGCAAGGAACTTCAAAAGATCATC 788
Db 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
QY 789 TGTGAACAAACGTACAGAAAAAATGGAACAAATGAAAAAGAGTTTGTGTACTGAAAAAG 848

Db 587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheCysValLeuLysLys 606
 QY 849 AAACCTGTGAGAGCAAAAGAAATATAATTCAGTTAGAGAACCAAAAGAGTAAATGGGAA 908
 Db 607 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValTrpGlu 626
 QY 909 CAAGAGCTCTCAGTGTGAGATTCACCTTAAACCAAGAGAGAGAGAGAGAGAGAGAG 968
 Db 627 GlnGluLeuLysSerValArgSerThrLeuAsnGlnGluGluGluLysArgArgAla 646
 QY 969 GATATATTAAATAGGAAAGAAATAGGAAAGAAATAGGAAAGAAATAGGAAAGAAATAG 1028
 Db 647 AsnLeuAsnGluLysLeuArgGluGluLysArgGluGluGlnHisArgLys 666
 QY 1029 GAGTTAGAGTGAACCAACCTTGAACAGCTCTCAGATACAGATACAGATACAGATACAG 1088
 Db 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgGlnAspLeuGluLys 686
 QY 1089 AGTGTAGAAAGTAAATTTGAATCAGGTTTCTCAGCTCATGAAATGAAATATCTCTTA 1148
 Db 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
 QY 1149 CATGAATATTCAGTTCAAAAGGAAATTCGCAATGCTAAAGCTGAAATAGCCACCTG 1208
 Db 707 HisGluAsnCysMetLeuLysGluLysGluLeuAlaMetLeuLysGluLysLeuAlaThrLeu 726
 QY 1209 AAACCAATACAGGAAAGGAAATATAATCTTTGAGGACATTAAGATTTTAAAGAA 1268
 Db 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspLeuLysLeuLysGlu 746
 QY 1269 AAGATGCTGAGCTGAGACACCTAAAGCTGAAAGAGAGAAATCAATTAAGAGGCA 1328
 Db 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLysLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGGAGCTTAAAGTTCTGATAGCTGAGAACCAATGCTCACTTCTAA 1388
 Db 767 SerGlnTyrSerGlyGlnLeuLysValLeuLeuAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAGGAAACAGCAAGAAATAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1448
 Db 787 LeuLysGluLysGlnAspLysGluLeuGluAlaGluLeuGluLysSerHisProArg 806
 QY 1449 CTGGCTTCTGCTGACAGACCATGATCAATTTGTGACATCAAGAGAGGAGGAGGAGGAG 1508
 Db 807 LeuAlaSerAlaValGlnAspHisAspGlnLeuValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTCCATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1568
 Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATCAATGAGGTGCTCCATCAACCACTTTCTGAAGCTCAAGAGGAGGAGGAG 1628
 Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
 QY 1629 AAGCTTAAATTAATCTCAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1688
 Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuAspGluAsnThrLeuValSer 886
 QY 1689 AGGAACATGCAAGAGAGCAACGCTGAAACACAGTGTCAATGAAAGGAGGAGGAGGAGGAG 1748
 Db 886 r-GluHisAlaGlnArgAspGluArgGluThrGlnCysGlnMetLysGluAlaGluHis 906
 QY 1749 TGTWTCAANGCAACAGATNATGTGACAAACACACTGACAGGAGGAGGAGGAGGAGGAGGAG 1808
 Db 906 eTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnGlnGlnGln 926
 QY 1809 AGAAATATTTCACACTCAAGCAAAATATGCGCTTCAACGAGCAATTAGTTCATGCA 1868
 Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnGlnGlnGlnGln 946
 QY 1869 ATAANGAAGCTGACAAACCAACAGATACCAATGATNTTCATTTCTTCAGAGGAGGAG 1928
 Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965

QY 1929 ATGCN-CATCATCTTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1987
 Db 966 MetGlnHisLysLeuLysGluLysAsnGluLysLysAsnTyrAsnAsnHisLeu 985
 QY 1986 AAAAACCCGATATTTCAATATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2027
 Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998
 RESULT 10
 US-09-551-621-475
 ; Sequence 475, Application US/09551621
 ; Publication No. US20030104366A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yuqun, Jiang
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Harlocker, Susan L.
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
 ; FILE OF INVENTION: DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
 ; FILE REFERENCE: 210121.470C5
 ; CURRENT APPLICATION NUMBER: US/09/551,621
 ; CURRENT FILING DATE: 2000-04-17
 ; NUMBER OF SEQ ID NOS: 479
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 475
 ; LENGTH: 1002
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 ; FEATURE:
 ; NAME/KEY: VARIANT
 ; LOCATION: (1)...(1002)
 ; OTHER INFORMATION: Xaa = Any Amino Acid
 US-09-551-621-475
 Alignment Scores:
 Pred. No.: 2,78e-232 Length: 1002
 Score: 3261.00 Matches: 654
 Percent Similarity: 97.19% Conservatives: 3
 Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: Gaps: 0
 US-09-602-362B-15 (1-2030) x US-09-551-621-475 (1-1002)
 QY 9 GTTAAAGATGCTTCTGAGGCTAACTCGCGAATGAAAGTTCTTATCCCACTAAAGCC 68
 Db 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCTCCCGAAGAGCCATTCGCTTC 128
 Db 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 366
 QY 129 GAGCTGCGCATTCGAAATGCAAGAGCTGTTCCAATTAAGCTTGGATTTGAAGATCAA 188
 Db 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAACATTTGAGACGACATGAGATCTCCCATCAGAAATCCAAACAAAGAGATTAAGAA 248
 Db 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGGGATTCAGAGTCTGTGAGACTCTTCAGAGAGGAGTGTGTGTTTACCC 308
 Db 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGCTTACATCAAAAG 368
 Db 427 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
 QY 369 AATGATGGTTTCTGAGGCTCCCTGCGAGATGAAGTTTCTTATCCCAACTAAAGCCCTTA 428
 Db 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466

429	QY	GAATTGATGGACATGCAAACTTTCAAAGCAGAGCCTCCCGAAGACCATCTGCCTTCGAG	488
430	QY		
431	QY		
467	Db	GlulLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu	486
432	QY		
433	QY		
489	QY	CTTGCCATTCAAAATGCAAAAGTCTGTCCAATAAAGCCTTGGAAATTTGAAGAATGAACAA	548
490	QY		
487	Db	ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln	506
491	QY		
492	QY		
549	QY	ACATTGAGACGATCAGATGTTCCTCTCAGAAATCAAAACAAAAGAGGTTGAAGAAAT	608
550	QY		
507	Db	ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn	526
551	QY		
609	QY	TCCTGGATTCTCGAGTCTCCGTGAGCTGTTTCACAGAAGGATGCTGTGACCAACAG	668
610	QY		
527	Db	SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys	546
611	QY		
612	QY		
669	QY	GCTACACATCAAAAGAAATGGATAAAATTAAGTGAAAATTTAGAAGATTCAACTAGCCTTA	728
670	QY		
547	Db	AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu	566
671	QY		
729	QY	TCAAAATCTTGATACAGTTCATCTCTGGAAGCAGCAAGGAACTTCAAAAGATCAC	788
730	QY		
567	Db	SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis	586
731	QY		
732	QY		
789	QY	TGTCAACACGCTACAGAAAATGGACAAATGCAAAAAGAAAGTTTGTGTACTCAAAAAG	848
790	QY		
587	Db	CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys	606
791	QY		
849	QY	AAACTGTCAAGACCAAGAAATAAATCACAGTTTAGAGAACCAAGCAAAAGTTAAATCGGAA	908
850	QY		
607	Db	LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu	626
851	QY		
852	QY		
909	QY	CAAGAGCTCTGCAGTGTGAGATTGACTTTAAACCAAGAAGAGAGAGAGAAATGCC	968
910	QY		
627	Db	GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla	646
911	QY		
969	QY	GATATATTAATGAAAATATTGGAGAGAAATTAGAGAAGATTCGAGAGCAGCATCGAAA	1028
970	QY		
647	Db	AspIleLeuAsnGlnLysIleArgGluGluLeuGlyArgIleGluGlnGlnHisArgLys	666
971	QY		
1029	QY	GAGTTAGAAGTGAACAAACAACTTGCAACAGGCTCTCAGAATACAGAATATAGAAATCAAG	1088
1030	QY		
667	Db	GluLeuGluValLysGlnGlnLeuGlnAlaLeuArgIleGlnAspIleGluLeuLys	686
1031	QY		
1089	QY	AGTGTAGAAATTAATTGAATCAGGTTTCTCACACTCATGAATGAATTAATTCCTTA	1148
1090	QY		
687	Db	SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnThrLeuLeu	706
1091	QY		
1149	QY	CATCAAAATTCGATGTGAAAAGGAAATGCCCATGCTTAAACTCGAAATAGCCACACTG	1208
1150	QY		
707	Db	HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu	726
1151	QY		
1209	QY	AAACACCAATACAGGAAAAGGAAAATAAATCTTTAGGACATTAAGATTTTAAAGAA	1268
1210	QY		
727	Db	LysHisGlnThrGlnGluLysGluAsnLysThrPheGluAspIleLysIleLeuLysGlu	746
1211	QY		
1269	QY	AAGAATCTCAACTTCAGATGACCCATAAACTGAAGAAGGAATCATTAACATAAAGGCCA	1328
1270	QY		
747	Db	LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluLysLeuThrLysArgAla	766
1271	QY		
1329	QY	TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACAAATGCTCACTTAA	1388
1330	QY		
767	Db	SerGlnThrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys	786
1331	QY		
1389	QY	TTGAAGGAAAAACAAGCAAGAAATFACTAGAGGCAGAAATTAATCACACCATCTTAGA	1448
1390	QY		
787	Db	LeuLysGlnLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg	806
1391	QY		
1449	QY	CTGCGTCTGCTGACCAAGACCATGATCAAAATTGTGACATCAAGAAAAGTCAAGACCT	1508
1450	QY		
807	Db	LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro	826

1509	QY	GC	TTTCCACATTCGAGAGATGCTGTTTGGCAAGAAAATGAATGTTGATGAGTAGT	1568
827	Db	Ala	PheHisIleIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer	846
1569	QY	ACC	GATATATAACAATGAGTGCTCATCAACCACTTTCTGAAGCTCAAAAGGAAATCCAN	1628
847	Db	Thr	IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLy	866
1629	QY	AAG	CTTAAAAATTAATCTCAATTATGCGAGGATGCTCTTAAGAGAAAATACATTGGTTTC	1688
866	Db	Ser	LeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgLysAsnThrLeuValSe	886
1689	QY	AGG	AACATGCAACAAAGACCAACGCTGAAACACACAGTGTCAATGAAGAGCACTGGAACACA	1748
886	Db	r-Glu	HisAlaGlnArgGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM	906
1749	QY	TG	TNTCAAACGCAAGATNATGTGACAAACACACTGANCACAGAGAGTCTCTAGATC	1808
906	Db	etTyr	GlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG	926
1809	QY	AGAA	TTATTTCACCTACAAAGCAAAAATATGTGCTTCAACAGCAATTAGTTTCATGCAC	1868
926	Db	lnLys	LeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisLah	946
1869	QY	ATA	ANGAAGCTGACCAACAAAAGCAAGATPACAAATTGANTTCTATTCTTGAGAGAAA	1928
946	Db	is-Lys	IysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys	965
1929	QY	ATGCN	-CATCATCTCTTAAAGAGAAAATGAGAGATATTTNATTACNATTAACCACTTTA	1987
966	Db	Met	GlnHisHisIleuLeuLysGlnLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu	985
1988	QY	AAAA	CCCGTATATTTCAATATGGAAGAAAANAAAAA	2027
986	Db	LysAsn	-ArgIleTyrGlnTyrGluLysGluLysAlaGlu	998

RESULT 11

```

US-10-007-805-475
: Sequence 475, Application US/10007805
: Publication No. US20020150581A1
: GENERAL INFORMATION:
: APPLICANT: Jiang, Yugu
: APPLICANT: Mitcham, Jennifer L.
: APPLICANT: Xu, Jiangchun
: APPLICANT: Harlocker, Susan L.
: APPLICANT: Hepler, William T.
: APPLICANT: Henderson, Robert A.
: APPLICANT: Farger, Gary R.
: APPLICANT: Vedvick, Thomas S.
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Durham, Margarita
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
: FILE REFERENCE: 210121.470C10
: CURRENT APPLICATION NUMBER: US/10/007,805
: NUMBER OF SEQ ID NOS: 593
: CURRENT FILING DATE: 2001-12-07
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 475
: LENGTH: 1002
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: VARIANT
: LOCATION: 310..429, 522
: OTHER INFORMATION: Xaa = Any Amino Acid
US-10-007-805-475

Alignment Scores:
Pred. No.: 2,78e-232 Length: 1002
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3

```

Best Local Similarity: 96.75% Mismatches: 15
 Query Match: 95.83% Indels: 5
 DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-475 (1-1002)

QY 9 GTTAAAGATGCTCTCTGAAGCTTAACCTGCGGATGAAGTTCTATTCCAACTAAGCC 68
 DB 327 VallysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
 QY 69 TTAGAATTGATGGACATCAAACTTTCAGACGAGCTCCCGAGAGCCCATCTGCCTTC 128
 DB 347 LeuGluLeuMetAspMetGlnThrPhelLysAlaGluProGluLysProSerAlaPhe 366
 QY 129 GAGCTGCGCATTAAGTAAAGCTGCTCCCAATTAAGCTTGGATTAAGAGATGA 188
 DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
 QY 189 CAACATTGAGACGATGAGATCTCCATCAAGATCCAAATCCAAAGAGCTATGAAGAA 248
 DB 387 GlnThrLeuArgAlaAspGluLeuLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
 QY 249 AGTTCTTGGATCTGAGAGCTCTGTGAGACTGTTTCAGAGAGGATGCTGTTTACC 308
 DB 407 SerSerIlePaspSerGluSerLeuLysGluThrValSerGlnLysAspValCysLeuPro 426
 QY 309 AAGCTTACACATCAAAAGAAATAGATAAAATAAATCGAAATTTAGAGAGTCTCTGAT 368
 DB 427 LysAla**HisGlnLysGluLeuLeuAspLysIleAsnGlyLysLeuGluLysProAsp 446
 QY 369 AATGATGTTCTTGAAGCTCCCTGAGATGAAGTTCTATTCCAACTAAGCTTA 428
 DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
 QY 429 GAATTGATGGATCGCAAACTTCAAGCAGAGCTCCCGAGAGCCATCTGCCTTCGAG 488
 DB 467 GluLeuMetAspMetGlnThrPhelLysAlaGluProGluLysProSerAlaPheGlu 486
 QY 489 CCTGCGCATTAAGTCAAAAGCTGTTCCCAATTAAGCTTGAATTTGAAGATGAACA 548
 DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLysAsnGluGln 506
 QY 549 ACATTGAGACGATCGATCTCCCTCAAGATCAAAATCAAAAGAGGTTGAAGAAAT 608
 DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
 QY 609 TCTTGGGATCTGAGAGCTCCGTCGAGACTGTTTCAGAGAGGATGCTGTGTACCCAG 668
 DB 527 SerIlePaspSerGluSerLeuLysGluThrValSerGlnLysAspValCysValProLys 546
 QY 669 GCTACATCAAAAGAAATCGATAAATAAGTGAAGAAATTAGAAGATCAACTAGCCTA 728
 DB 547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluLysSerThrSerLeu 566
 QY 729 TCAAAATCTTGGATACAGTTCTTCTGTGAAGACGACAGGCACTTCAAAAGATCAC 788
 DB 567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuLysAspHis 586
 QY 789 TGTGAACACTACAGGAAATCGAACAATGAAGAAATGAAGTTTGTGTACTGAAAGAG 848
 DB 587 CysGluGlnArgThrGlyLysMetGluLysMetLysLysLysPheCysValLeuLysLys 606
 QY 849 AAATCTGTCAGAGCAAAAGAAATTAATCAAGTTAGAGAACCAAAAGTTAAATGGGAA 908
 DB 607 LysLeuSerGluAlaLysGluLysSerGlnLeuGluAsnGlnLysValLysTrpGlu 626
 QY 909 CAAGAGCTCTCGAGTGTGAGATTCGATTTTAAACCAAGAGAGAGAGAGAGAAATGCC 968
 DB 627 GlnGluLeuCysSerValArgLeuThrLeuAspGlnGluLysArgAlaAsnAla 646
 QY 969 GATATTTAAATGAAATTTAGGAGAAATTTAGGAGAAATTCAGAGAGAGATAGGAA 1028
 DB 647 AspIleLeuAsnGluLysIleArgGluLeuGlyArgIleGluGlnHisArgLys 666

QY 1029 GAGTTGAAGTGAACCAACCAACTTGAACAGGCTCTCAGAAATCAAGATATAGAAATGAAG 1088
 DB 667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
 QY 1089 AGTGTCAAAAGTAAATTTGAATCAGGTTCTTCACACTCATGAATAAGAAATATCTCTTA 1148
 DB 687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
 QY 1149 CATCAAAATTCGATGTGTTGAAAAAGGAAATTCGATCTAAAACTGGAATAGCCACTG 1208
 DB 707 HisGluAsnCysMetLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
 QY 1209 AAACACCAATACAGAGAAAGGAAATAAATACTTTGAGCAGATTAAGATTTTAAAGAA 1268
 DB 727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
 QY 1269 AAGATCTCTGAATCTCAGATGACCTTAAACTGAAAGAGGAATCAATTAACATAAGGCA 1328
 DB 747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
 QY 1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACCAATGCTCACTCTAAA 1388
 DB 767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
 QY 1389 TTGAAGGAAAAACAAGCAAAAGAAATCTAGAGGCGGAAATTTGAATCAACATCTCTAGA 1448
 DB 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisProArg 806
 QY 1449 CTGCTTCTCTCTGTACAGACCATGATCAATTTGTGACATCAAGAAAAAGTCAAGAACT 1508
 DB 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
 QY 1509 GCTTTCACATCTCCAGAGATGCTTGTGTTGAAAGAAAAATGAATGTTGATGAGTAGT 1568
 DB 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
 QY 1569 ACCGATATATAACAAATGAGTGTCTCCATCAACACTTCTGAAGCTCAAGGAAATCCAN 1628
 DB 847 ThrIleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
 QY 1629 AAGCTTAAATTAATCTCAATTTATGACAGAGATGCTCTAAGAGAAATATACATTCGTTTC 1688
 DB 866 SerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 886
 QY 1689 AGGAACATGCACAAAGAGACCAACGTAACAGTGAACACAGTGTCAATGAAGAAAGTCAACACA 1748
 DB 886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
 QY 1749 TGTNTCAANCGAAACAAGATNATGTGAACAAACACACTGAGCAGAGAGTCTTAGATC 1808
 DB 906 etTyrGlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGlnSerLeuAspG 926
 QY 1809 AGAATTTATTTCACTACAAAGCAAAATATGCTGCTTCAACAGCAATTAGTTCATGCAC 1868
 DB 926 LysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
 QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATACAAATGTATTTTCATTTCTTTCAGAGGAA 1928
 DB 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
 QY 1929 ATGCTCATCTCTTAAAGAGAAAAATAGAGAGATTTTATTTATTCATTAATCAATTTA 1987
 DB 966 MetGlnHisHisLeuLeuLysGluLysAsnGluIlePheAsnTyrAsnAsnHisLeu 985
 QY 1988 AAAAAACCTATTTTCAATATTCGAAAAAANAANAANAANAANAANAANAANAANAANA 2027
 DB 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 12
 US-10-076-622-475
 ; Sequence 475, Application US/10075622
 ; Publication No. US20030023036A1

```
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 475
; LENGTH: 1002
; TYPE: PRN
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 310, 429, 522
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-076-622-475

Alignment Scores:
Pred. No.: 2,78e-232 Length: 1002
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 14 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-076-622-475 (1-1002)
QY 9 GTTAAAGATGGTCTTCTGAAGGCTAACTGGGAATGAAAGTTTCTATTCCAACTAAAGCC 68
DB 327 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 346
QY 69 TTAGAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAAGCCATCTCCCTTC 128
DB 347 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPhe 366
QY 129 GAGCCTGCCATTGAAATGCAAAAGTCTGTCCAAATAAAGCCTTGGAAATGGAAGATGAA 188
DB 367 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 386
QY 189 CAAACATTGAGAGCAGATGAGATCTCCCATCAGAAATCCAAACAAAAGGACTATGAAGAA 248
DB 387 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 406
QY 249 AGTTCTTGGATCTCAGAGTCTCTGTGAGACTGTTTCACAGAGGATGTGTGTTACCC 308
DB 407 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 426
QY 309 AAGGCTACACATCAAAAAGAAATAGATAAAATAAATGGAATTAAGAGAGTCTCTCAT 368
DB 427 LysAla**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 446
QY 369 AATGATGTTTCTGAAGCTCCCTGCGAATGAAAGTTTCTATTCCAACTAAAGCCTTA 428
DB 447 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 466
QY 429 GAATTGATGACATGCAAACTTTCAAAGCAGAGCCTCCCGAAGCCATCTGCTTCAG 488
DB 467 GluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAlaPheGlu 486
QY 489 CTGCGCATTAATGCAAAAGTCTTCCAAATAAAGCCTTGGAAATGGAAGATGAACAA 548
DB 487 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 506
QY 549 ACATTGAGAGCAGATGATGCTTCCCAATCAAAATCAAAACAAAAGAGGTTGAAGAAAT 608
DB 507 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 526
QY 609 TCTTGGGATTTCTGAGAGTCTCGTGAGACTGTTTCACAGAGGATGTGTGTACCCAAG 668
DB 668 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisn 906

527 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 546
669 GCTACACATCAAAAAGAAATCGATAAAATAAGTGGAAAATTAGAGATTCACCTAGCCTA 728
547 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 566
729 TCAAAAATCTTTGGATACAGATTCTTCTTTGTAAGAGCAAGGAACTTCAAAAAGATCAC 788
567 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 586
789 TGTGACACAGCTACAGGAAAATGGAACAATGAAAAGAAAGTTTGTGTCTCTGAAAAG 848
587 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 606
849 AAACCTGTCAGAAAGCAAAAGAAATAAATCAAGTTAGAGAACCAAAAAGTTTAAATGGCAA 908
607 LysLeuSerGluAlaLysGluIleLysSerGlnLeuGluAsnGlnLysValIleTrpGlu 626
909 CAAGAGCTCTCGAGTGTGAGATTGACTTTAAACAAGAGAGAGAGAGAGAGAAATGCC 968
627 GlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGluLysArgArgAsnAla 646
969 GATATATTAAATGAAAATTTAGGGAAGAAATTAGGAAGAATCGAAGAGCAGCATAGAAA 1028
647 AspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys 666
1029 GAGTTAGAAGTCAAAACAACTTTCAACAGGCTCTCAGAATACAGAATATAGATTGAAG 1088
667 GluLeuGluValLysGlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys 686
1089 AGTCTAGAAATTAATTTGAATCAGTGTCTCAGCTCATCTCATGAAATGAAAATATCTCTTA 1148
687 SerValGluSerAsnLeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu 706
1149 CATGAAATTCATGTTTCAAAAAGGAAATGCCATGCTTAAACTGGAATAAGCACACTG 1208
707 HisGluAsnCysMetLeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu 726
1209 AAACACCAATACCAAGGAAAAGGAAAATAAATACCTTTGAGACATTAAGATTAAAAAGAA 1268
727 LysHisGlnTyrGlnGluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu 746
1269 AAGAATCTCAACTTCAGATGACCTTAAACTCAAAAGAGCAATCATTAACATAAAGGCA 1328
747 LysAsnAlaGluLeuGlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla 766
1329 TCTCAATATAGTGGCAGCTTAAAGTTCTGATAGCTGAGAACACAATGCTCACTTCTAAA 1388
767 SerGlnTyrSerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
1389 TTGAAGGAAAAACAAGACAAAGAAATAGTAGGCGAGAAATTCGAATCAACACCATCTCTAGA 1448
787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisSHISProArg 806
1449 CTGCTTCTGCTGTACAAGACCATGATCAAAATGTCACATCAAGAAAAAGTCAAGAACCT 1508
807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
1509 GCTTTCCACATTCAGAGAGATGCTTCTTCCAAAGAAAATGNAATGTTGATGTGAGTAGT 1568
827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
1569 ACCGATATATAACAATGAGGTGCTCCATCAACACCTTTCTGAAGCTCAAGGAAATCCAN 1628
847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGluArgLysSerLys 866
1629 AACCTTAAAAATTAATCTCAATTATCCAGAGATGCTCTAAGAGAAAAATACATTGTTTC 1688
866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValSer 886
1689 AGAAACATGCACAAAGAGACCAACGTGAACACAGTGTCAATGAGGAGGCTGACACACA 1748
886 r-GluHisAlaGlnArgAspGlnArgGluThrGlnCysGlnMetLysGluAlaGluHisn 906
```



```
Db 767 SerGlnTyr:SerGlyGlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys 786
QY 1389 TTGAAGGAAACACAGACAAAGAAATACCTAGAGGCGAGAAATGAATCAACACATCCTAGA 1448
Db 787 LeuLysGluLysGlnAspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg 806
QY 1449 CTGGCTTCTGCTGACACACATGATCAAAATTTGTGACATCAAGAAAGTCAAGAACT 1508
Db 807 LeuAlaSerAlaValGlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro 826
QY 1509 GCTTTCCACATTCGAGGAGATCTGTGTTTGCAGAGAAATCAATGTTGATGTAGTAGT 1568
Db 827 AlaPheHisIleAlaGlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer 846
QY 1569 ACCGATATATAACATGAGTCTCCATCAACACCTTTCTGAGCTCAAGAGAAATCCAN 1628
Db 847 Thr-IleTyrAsnAsnGluValLeuHisGlnProLeuSerGluAlaGlnArgLysSerLys 866
QY 1629 AAGCCTAAAATTAATCTCAATATCCAGAGATGCTCTAAGAGAGAAATACATTTGTTTC 1688
Db 866 sSerLeuLysIleAsnLeuAsnTyrAlaGlyAspAlaLeuArgGluAsnThrLeuValIse 886
QY 1689 AGGACATGCACAAAGACACCACTGAAACACAGTGTCAATGAAGAGAGCTGAACACA 1748
Db 886 r-GluHisAlaGlnArgAspGlnA:rgGluThrGlnCysGlnMetLysGluAlaGluHisM 906
QY 1749 TCTNTCAAAACCAAGCATNATGTCAACAAACACACTGANCAGCAGAGTCTCTAGATC 1808
Db 906 etTyr-GlnAsnGluGlnAspAsnValAsnLysHisThrGluGlnGlnGluSerLeuAspG 926
QY 1809 AGAATATTATCACTACAAAGCAAAATATATGTGCTTCAACAGCAATAGTTCATGCAAC 1868
Db 926 InLysLeuPheGlnLeuGlnSerLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaH 946
QY 1869 ATAANGAAGCTGACAAACAAAGCAAGATAACAATTTGATNTTTCATNTTCTTGAGAGAAA 1928
Db 946 is-LysLysAlaAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLys 965
QY 1929 ATGCN-CATCAPCTTCTAAAGAGAGAAATAGGAGATATTNNATACNATAACCATTTA 1987
Db 966 MetGlnHisHisLeuLysGluLysAsnGluGluIlePheAsnTyrAsnAsnHisLeu 985
QY 1988 AAARACCCGTATATTCAATATGGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2027
Db 986 LysAsn-ArgIleTyrGlnTyrGluLysGluLysAlaGlu 998

RESULT 14
US-10-007-805-493
; Sequence 493, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Xuqiu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 493
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 403, 522, 615
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-007-805-493

Alignment Scores:
Pred. No.: 2,81e-232 Length: 1095
Score: 3261.00 Matches: 654
Percent Similarity: 97.19% Conservative: 3
Best Local Similarity: 96.75% Mismatches: 15
Query Match: 95.83% Indels: 5
DB: 13 Gaps: 0

US-09-602-362E-15 (1-2030) x US-10-007-805-493 (1-1095)
QY 9 GTTAAAGATGCTCTTCTGAAGGCTAACTGCGGATGAAGTTTCTATTCCAACTAAAGCC 68
Db 420 ValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLysAla 439
QY 69 TTGAATTTGATGACATGCAAACTTTTCAAGCAGAGAGCTCCCGAGAGAGCCATCTGCCTTC 128
Db 440 LeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPhe 459
QY 129 GAGCCTCCCATTTGAATGCAAAAGTCTGTTCCTCAAAATAAGCCTTGGAAATTTGAAGATGAA 188
Db 460 GluProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu 479
QY 189 CAACATTTGAGAGCAGATGAGATACCTCCATCAGATCCAAACAAAGAGACTATGAAGAA 248
Db 480 GlnThrLeuArgAlaAspGluIleLeuProSerGluSerLysGlnLysAspTyrGluGlu 499
QY 249 AGTTCTTTGGGATCTGAGAGTCTCTGTGAGACTGTTCACAGAGAGGATGTGTGTTTACCC 308
Db 500 SerSerTrpAspSerGluSerLeuCysGluThrValSerGlnLysAspValCysLeuPro 519
QY 309 AAGCTACACATCAAAAGAAATAGATATAAATAATGGAATAATAGAGAGTCTCTCTGAT 368
Db 520 LysAlaA**HisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSerProAsp 539
QY 369 AATGATGCTTTCTGAGGCTCCCTGCAGATCAAAAGTTTCTATTCCAACTAAAGCCTTA 428
Db 540 AsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLysAlaLeu 559
QY 429 GAATTTGATGACATGCAAACTTTCAAAGCAGAGAGCTCCCGAGAGAGCCATCTGCCTTCGAG 488
Db 560 GluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAlaPheGlu 579
QY 489 CTTGCCATTTGAAATGCAAAAGTCTGTTCCTCAAAATAAGCCTTGAATTTGAAGATGACAA 548
Db 580 ProAlaIleGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGluGln 599
QY 549 ACATTGAGAGCAGATCAGATGTTCCCTTTCAGAAATCAAAACAAAGAGGTTGAAGAAAT 608
Db 600 ThrLeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys**ValGluGluAsn 619
QY 609 TCTTGGGATTTCTGAGAGTCTCCGTGAGACTGTTTTCAGAGAGATGTGTGTACCAAG 668
Db 620 SerTrpAspSerGluSerLeuArgGluThrValSerGlnLysAspValCysValProLys 639
QY 669 GCTACACATCAAAAGAAATGGAATAAATAAGTGAAGAAATTTAGAGATTTCAACTAGCCTA 728
Db 640 AlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu 659
QY 729 TCAAAATTTCTGGATTCAGTTCAATTTCTGTGAAGAGCAAGGAACTTCAAAAGATCATC 788
Db 660 SerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis 679
QY 789 TGTGAACACGTTACAGGAAATTTGGACAAATTCAAAGAAAGATTTTGTCTACTGAAAGAG 848
Db 680 CysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys 699
QY 849 AAATCTGTCAGAGCAAAAGAAATATAATACAGTTTAGAGAACCAAAAGTTTAAATGGGAA 908
```


Qy	489	CCTGCCATTGAAATGC	AAAACTGTTCTCCAAATAAAGCCTTTGGAAATTGAAGAATGAACAA	548
Db	580	ProAlaIleGluMetGln	LySerValProAsnLysAlaLeuGluLeuLysAsnGluGln	599
Qy	549	ACATTGAGACGACAT	CAGATGTCCTCCCTTCAGAATCAAAACAAAAGAAGGTTGAAGAAAT	608
Db	600	ThrLeuArgAlaAsp	GlnMetPheProSerGluSerLysGlnLys**ValGluGluLysn	619
Qy	609	TCTTGGATTCGTAGAG	TCTCCGTGAGACTGTTTCACAGAAGGATGTGTGTGTAACCCAAAG	668
Db	620	SerTrpAspSerGlu	SerLeuArgGluThrValSerGlnLysAspValCysValProLys	639
Qy	669	GCTACACATCAAAAG	AAATCGATAAATAGTCGAAATTTAGAAGATTTCAACTAGCCTA	728
Db	640	AlaThrHisGlnLys	GluMetAspLysIleSerGlyLysLeuGluAspSerThrSerLeu	659
Qy	729	TCAAAAATCTTTGGAT	CATACAGTTTCATTTGTGAAAGACAAAGGGAACATTCAAAAGATCAC	788
Db	660	SerLysIleLeuAsp	ThrValHisSerCysGluArgAlaArgGluLeuGlnLysAspHis	679
Qy	789	TGTGAACACGTACAG	GAANAATCGAACAAATGAAAAAGAAGTTTTGTGTACTGAAAAAG	848
Db	680	CysGluGlnArgThr	GlyLysMetGluGlnMetLysLysLysPheCysValLeuLysLys	699
Qy	849	AAACTGTCAAGAACA	AAAGAAATAAATACAGTTAGAGAACCAAAAGATTAAATGGCAA	908
Db	700	LysLeuSerGluAla	LysGluIleLysSerGlnLeuGluAsnGlnLysValLysTrpGlu	719
Qy	909	CAAGAGCTGTCAGTG	TGAGATTGACTTTAAACCAAGAGAAGAGAAGAAATGCC	968
Db	720	GlnGluLeuCysSer	ValArgLeuThrLeuAsnGlnGluGluLysLysArgAsnAla	739
Qy	969	GATATATTAAATGAAA	AAAAATPAGGAAGAATTAGGAAGAATCGAAGACGACGATAGGAAA	1028
Db	740	AspIleLeuAsnGlu	LysIleArgGluGluLeuGlyArgIleGluGluGlnHisArgLys	759
Qy	1029	GAGTTAGAAGTGAACA	CAACTTGAACAGGCTCTCAGAATACAAGATATAGAAATTGAAG	1088
Db	760	GluLeuGluValLys	GlnGlnLeuGluGlnAlaLeuArgIleGlnAspIleGluLeuLys	779
Qy	1089	AGTGTAGAAGTAAT	TTGAATCAGGTTCTCACACTCATGAATGAATATATCTCTTA	1148
Db	780	SerValGluSerAsn	LeuAsnGlnValSerHisThrHisGluAsnGluAsnTyrLeuLeu	799
Qy	1149	CATGAAATTCGATGTT	GAAAAAGGAATTCGCAATGCTAAACTGGAATATAGCACACTG	1208
Db	800	HisGluAsnCysMet	LeuLysLysGluIleAlaMetLeuLysLeuGluIleAlaThrLeu	819
Qy	1209	AAACACCAATACAG	GAAGAAATAATACCTTTGAGACATTAAGATTTTAAAGAA	1268
Db	820	LysHisGlnTyrGln	GluLysGluAsnLysTyrPheGluAspIleLysIleLeuLysGlu	839
Qy	1269	AAGAATGCTGAAC	TTCAGTCACCTTAAACTGAAAGAGAATCATTAACCTAAAGGGCA	1328
Db	840	LysAsnAlaGluLeu	GlnMetThrLeuLysLeuLysGluGluSerLeuThrLysArgAla	859
Qy	1329	TCTCAATATAGTGG	CGAGCTTAAAGTTCTGTAGACTGAGAACCAATGCTCACTTCTTAA	1388
Db	860	SerGlnTyrSerGly	GlnLeuLysValLeuIleAlaGluAsnThrMetLeuThrSerLys	879
Qy	1389	TTGAAGGAAGAAACA	AGACAAAGAATACTAGAGCGACAAATTGAAATCACACCATCTCTAGA	1448
Db	880	LeuLysGluLysGln	AspLysGluIleLeuGluAlaGluIleGluSerHisHisProArg	899
Qy	1449	CTGCGCTTCTGCT	GTACAGACCATGATCAAAATTCGTGACATCAAGAAAAAGCTCAAGAACCT	1508
Db	900	LeuAlaSerAlaVal	GlnAspHisAspGlnIleValThrSerArgLysSerGlnGluPro	919
Qy	1509	GCTTTCCACATTC	GAGGATGCTTGTGTTGCAAGAAAAATGAATGTGTAGTGAAGTACT	1568
Db	920	AlaPheHisIleAla	GlyAspAlaCysLeuGlnArgLysMetAsnValAspValSerSer	939
Qy	1569	ACCGATATATAACA	ATAGAGGTGCTCCCATCAACCACTTTCTGAAGCTCAAAAGGAAATCCAN	1628

Search completed: July 15, 2004, 09:42:46
Job time : 159.008 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:46:47 ; Search time 193.218 Seconds
(without alignments)
13439.296 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086

Sequence: 1 ctactctatacagcaacgac.....gagtggaactccactggaaa 4115

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO spoop/US09602362/runat_15072004_093626_22001/app_query.fasta_1.10325
-DB=SPTREMBL_25 -QFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=spt -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09602362 @CGN 1.1 8.4 @runat_15072004_093626_22001 -ICPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_muc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_vodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1234	17.4	1341	4	Q9BXX3	Q9bxx3 homo sapien

ID	Q9BXX3	PRELIMINARY;	PRT;	1341 AA.
AC	Q9BXX3			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Breast cancer antigen NY-BR-1.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21174979; PubMed=11280766;			
RA	Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,			
RA	Jaeger E., Knuch A., Old L.J., Chen Y.F.;			
RT	"Identification of a Tissue-specific Putative Transcription Factor in			
RT	Breast Tissue by Serological Screening of a Breast Cancer Library."			
RL	Cancer Res. 61:2055-2061(2001).			
DR	EMBL: AF269087; AAK27325.1; -			
DR	GO: 0005634; C.nucleus; NAS.			
DR	GO: 0005515; F.protein binding; NAS.			
DR	GO: 0003700; F.transcription factor activity; NAS.			
DR	GO: 0006355; P.regulation of transcription, DNA-dependent; NAS.			

ALIGNMENTS

RESULT 1

Q9BXX3 ID Q9BXX3 PRELIMINARY; PRT; 1341 AA.

ID	Q9BXX3	PRELIMINARY;	PRT;	1341 AA.
AC	Q9BXX3			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)			
DE	Breast cancer antigen NY-BR-1.			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=21174979; PubMed=11280766;			
RA	Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,			
RA	Jaeger E., Knuch A., Old L.J., Chen Y.F.;			
RT	"Identification of a Tissue-specific Putative Transcription Factor in			
RT	Breast Tissue by Serological Screening of a Breast Cancer Library."			
RL	Cancer Res. 61:2055-2061(2001).			
DR	EMBL: AF269087; AAK27325.1; -			
DR	GO: 0005634; C.nucleus; NAS.			
DR	GO: 0005515; F.protein binding; NAS.			
DR	GO: 0003700; F.transcription factor activity; NAS.			
DR	GO: 0006355; P.regulation of transcription, DNA-dependent; NAS.			

```

DR InterPro; IPR002110; ANK.
DR InterPro; IPR001969; Asprotease_AS.
DR Pfam; PF00023; ANK; 6.
DR SMART; SM00248; ANK; 6.
DR PROSITE; PS00088; ANK REPEAT; 4.
DR PROSITE; PS00297; ANK REP REGION; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW ANK repeat; Repeat.
SQ SEQUENCE 1341 AA; 152777 MW; 33E53DDB6FD3A58B CRC64;

Alignment Scores:
Pred. No.: 4,42e-113 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 4 Gaps: 115

US-09-602-362E-22 (1-4115) x Q9BXX3 (1-1341)
QY 96 CAAGGAGAGAGACATCAACCTTAATATACAGAG---CCAGAGAGAGCTGCTCTAAC 152
DB 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGGTCAATGGCCT-----GAGGAAGTAGTAACATTC---GGTA 194
DB 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38
QY 195 GACAGAGTGCAGCTGAGCTCTG---TGGCGAACACGACACCTCTGATGAGGC-T 250
DB 38 sPArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAla 58
QY 251 TACAATGCCACAGGAGGCTTGC---AAATATTTGATAGATTC-GGTCCCATATAATC 306
DB 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGCTTCCATTATCGCTTTATAGTGGATTTGTCAGT- 362
DB 78 euValAspValTyrglyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 ---GGTGGCAACTGCTGTC-CATGCTGCAGTATCAAGTGC---CAACAGGCTGCCTCA 416
DB 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
QY 417 CACCACTTTTACTACCAT---AACGAAAGAGTAGCAATTGTGGAATTT---GCTGAT 470
DB 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAATGCAATGCGAATGCGTTA---ATAAGTTAAATGCACCCCTCATGCTGTAT 527
DB 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
QY 528 G---TCTGGATCATCAGATAGTTGTCATGCTTCTTAGCAAAA---TGTGACGTCCTT-- 579
DB 157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
QY 580 -GTGCAGATATAGTAGTAAGTAAGTACGAGACAT---ATGCTGTACTGTGGATTTCTACA 635
DB 177 alAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisI 197
QY 636 TTCATAACAAA---TTATGAATATATACGAAATATCAAAATATCA---TCAAAATACAA 689
DB 197 leHisGluGlnIleMet-GluTyrlleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAAGAACTCTGCAGAAACCTGATGAGCTGCACCT---GGCGGAAGACAC 743
DB 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
QY 744 CTGACACGCTGAAAGCT---GGTGCAAAAAACACCTGATGAGCTGCACCT---GGTGA 797
DB 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
QY 798 AAGACACCTGACAGGCTGAAAGCT---GGTGCAAAAAACACCTGATGAGGCTGCATCT- 851

```



```

QY 3769 ATT---CAATATGAAAGAGAA 3789
Db 1328 lIeTyrGlnTyrGluLysGluLys 1335

RESULT 2
ID Q9BXX2 PRELIMINARY; PRT; 1011 AA.
AC Q9BXX2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Breast cancer antigen NY-BR-1.1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21174979; PubMed=11280766;
RA Jaeger D., Stockert E., Guere A.O., Scanlan M.J., Karbach J.,
RA Jaeger E., Knuth A., Old L.J., Chen Y.T.;
RT "Identification of a Tissue-Specific Putative Transcription Factor in
RT Breast Tissue by Serological Screening of a Breast Cancer Library.";
RL Cancer Res. 61:2055-2061 (2001).
DR EMBL; AF269088; AAK27326.1; -.
DR InterPro; IPR002110; ANK.
DR Pfam; PF00023; ank; 3.
DR SMART; SM00248; ANK; 3.
DR PROSITE; PS00088; ANK_REPEAT; 1.
DR PROSITE; PS0297; ANK_REPEAT_REGION; 1.
KW ANK repeat; Repeat.
FT NON TER 1011 1011
SQ SEQUENCE 1011 AA; 114250 MW; 16627D7B218DE438 CRC64;

Alignment Scores:
Pred. No.: 2,34e-38 Length: 1011
Score: 481.00 Matches: 363
Percent Similarity: 44.39% Conservative: 135
Best Local Similarity: 32.35% Mismatches: 303
Query Match: 6.79% Indels: 327
DB: Gaps: 75

US-09-602-362E-22 (1-4115) x Q9BXX2 (1-1011)

QY 380 CATGGTCAGTATCGAAGTGCRAA---CAAGGCTGCCTCACACCTTTTACTACCAT-- 434
Db 8 TyrGlyAlaValIleGluValGlnAsnLysAlaSerLeuThrProLeuLeuLeu-Ala11 27
QY 435 -AACGAAAGAGTCGAGCAATTGGGAATTTT---GCTGATAAAATGCAATCGAATGC 490
Db 27 eGlnLysArgSerLysGlnThrValGluPheLeuLeuThrLysAlaAsnAlaAsnAl 47
QY 491 GTTA---ATAAGTTAAATGCACACCCCTCATGCTGCTGATG---TCTGGATCATCAGAGA 544
Db 47 aPheAsnGluSer-LysCysThrAlaLeuMetLeuAlaIleCysGluGlySerSerGluI 67
QY 545 TAGTTGTCATCTCTTAGCAAAA---TGTCACCTCTTTTGTCAG---ATATAGTGGAGTA 598
Db 67 leValGlyMetLeuLeuGlnGlnAsnValAspValPheAlaGluAspIle-HisGlyIle 86
QY 599 ACTGCAGAA---CATATGCTGTTTACTGTGATTCATCATTCATAACAAATTTATGAATA 655
Db 87 ThrAlaGluArgTyrAlaAlaAlaArgGlyValAsnTyrIleHisGlnGlnLeuLeuGlu 106
QY 656 TAT---ACGAAATATCAAAATCATCAATACCAATCAGAGGAACCTCTGCGAGAAAC 712
Db 107 HisIleArgLysLeuProLysAsn----- 114
QY 713 CTGATGAGGCTGCACCTCGCGGAAAGACACCTGCACACCTGAAAGCTGGTGGAAAAACA 772
Db 115 -----ProGlnAsnThrAsnProGluGlyThrSerThrGly-----Thr 127
QY 773 CTTGATGAGGCTGCACCT---GGTGGAAAGACACCTGCACGCTGAAAGCT---GGTGC 826

```

```

Db 128 ProAspGluAlaAla-ProLeuAlaGluArgThrProAspThrAlaGluSerLeuLeuG1 147
QY 827 AAAAACACCTGATGAG---GCTGCATCTTGGTGGAGAACATCTGCACAAATTCAT--G 880
Db 147 uLysThrProAspGluAlaAla-ArgLeuValGluGlyThrSerAlaLysIleGlnCysL 167
QY 881 TTGGAGAAAGGACATCTGAAAAGTTGGAACGTGAGC---AGAAAAACACCTAGGAAA-- 934
Db 167 euGlyLysAlaThrSerGlyLysPheGluGln-SerThrGluGluThrProArgLysIle 186
QY 935 TTACGATCTCGCAAAAACATCTGAGAAATTTACGGG-CCAGCAAGAGACACCTAGAA 993
Db 187 LeuArgProThrLysGluThrSerGluLysPheSerTrpProAlaLysGlu----- 203
QY 994 GATCGCAGGAGAGAAAAGAACACACCTAGGGAATATAGTCCCGAAAAAGAACACTG 1053
Db 204 -----ArgSerArgLysIleThr-----TrpGluGluLysGluThrSerVal 217
QY 1054 AGAATTCCTGGGAGAAAAGAACAGACCTAGGAGAGCATGGGAGAAAAGAACCCCT 1113
Db 218 LysThrGluCysValAlaGlyValThrProAsnLysThrGluValLeuGluLysGlyThr 237
QY 1114 GTAAAGATCGATCGCTGCAAGAGTAACATCTAATAAACTAAA----- 1155
Db 238 SerAsnMetIleAlaCysProThrLysGluThrSerThrLysAlaSerThrAsnValAsp 257
QY 1155 ----- 1155
Db 258 ValSerSerValGluProIlePheSerLeuPheGlyThrArgThrIleGluAsnSerGln 277
QY 1156 -----GTTTCAAAAAGCAATC 1173
Db 278 CysThrLysValGluGluAspPheAsnLeuAlaThrLysIleIleSerLysSerAlaAla 297
QY 1174 TAAGATGTTGCATGT---CCACAAAAGAAATCAT-----CTACAAAGC 1212
Db 298 GlnAsnTyrThrCysLeuProAspAlaThrTyrGlnLysAspIleLysThrIleAsnHis 317
QY 1213 AAGTCCCATGATCAGAGG---TCCCATCAGATCCAAACAGGAGATGAGATATCT 1269
Db 318 LysIleGluAspGlnMetPheProSerGluSerLysArgGluGluAspGluGlyThrSer 337
QY 1270 TTGATTCTCGGCTCTCTTTGAAGTTCTGC-----AAGATTCAAGTG---GTATACCTCAGT 1322
Db 338 TrpAspSerGlySerLeuPheGluSerSerAlaLysThrGlnValCysIleProGluSer 357
QY 1323 CTATATATAAAGTAATGATGATATAAAGAG---TAGAAGCCTCTCTAAAGCCATCT 1379
Db 358 MetTyrGlnLysValMetGluIleAsnArgGluValGluGluLeuProGluLysProSer 377
QY 1380 GCCTTCAAGCCGCCATGTAAA---GCCAAACTCTTCCAAATAAGCCT---TTGAATGAA 1433
Db 378 AlaPheLysProAlaValGluMetGlnLysThrValProAsnLysAlaPheGlu-LeuL 397
QY 1434 GAATGAACAAACATTAAGACAGATCCGTGTT---CCACCCGAATCCAAACAAAGG---AC 1487
Db 397 sAsnGluGlnThrLeuArgAlaAlaGlnMetPheProSerGluSerLysGlnLysAspAs 417
QY 1488 TATAAGAAAATTTCTGGATTCTAGAGTCT---CTGGAGACTGTTTACAGAGGATTTGTG 1544
Db 417 pGlu-GluAsnSerTrpAspSer-GluSerProCysGluThrValSerGlnLysAspVal 436
QY 1545 TTT---ACCAGGCTACACATCAAAAAGAA---TAGATAAAAAAATGCAAAATAGAGAG 1598
Db 437 TyrLeuProLysAlaThrHisGlnLysGluPheAspThrLeuSerGlyLysLeuGluGlu 456
QY 1599 TCCCTATAAAGTGGTCTTCTGAGAGGCTACCTCGG-----AATGAAATTTCT 1645
Db 457 SerProValLys-----AspGlyLeuLeuLysProThrCysGlyArgLysValSer 473
QY 1646 ATTCCACTAAGCCT---AGAATTTGAGACATCAAACTTTTCAAGCG----- 1690

```


Percent Similarity:	37.01%	Conservative:	173
Best Local Similarity:	20.84%	Mismatches:	393
Query Match:	2.17%	Indels:	284
DB:	5	Gaps:	43
US-09-602-362E-22 (1-4115) x Q8I659 (1-1383)			
QY	964	ATTTCAGGGCCAGCAAGAGAGACCTAGAGATCGCAGGGAGAAAAAGACACACCTA	1023
DB	3	IleLeuGlyAspGluLysGluSerTyrThrSerTyrCysAspLysAsnAsnAspIleLeu	22
QY	1024	-----GGGAATAATGAGTCCTCCGAAAAAGAAACACTGAGAAATTAGCTGGGAGAAAA	1074
DB	23	IleSerLysGlnLysIleGluAspLeuLysLysSerIleGluAsnLeuLeuAsnAspLys	42
QY	1075	GGAAGACCTAGAGACCGCATGGGAGAAAAAGAAACCTGTAAAGATGGATGGCTGCAAG	1134
DB	43	AsnAlaHisTyrGluLeuAsnHisIleLysArgSerLeuAsnGluLeuAspIleTyrThr	62
QY	1135	AGTACATCTAATAACTAAAGTTTGTGAAAAAGGAAATCTAAGATGTTGCATGTCCACAA	1194
DB	63	LysAsn-----LysSerAspLeuPheAsn-----	70
QY	1195	AAGATCATCTCAAGAACGATGCGATCATCAGAGTCCCATCAGATCCAAACAGGGAA	1254
DB	71	AsnTyrAsnLeuAsnGluSerLeuAsnAsp-----SerTyrLysAspLysThrPheGlu	88
QY	1255	GATGAAGATATCTTTCATCTCGG-----GTCCTCTTCAAGTTC	1295
DB	89	GluLeuLysSerGlnLeuLysGlnLysAsnLeuAsnSerCysLeuSerLeuLysLeu	108
QY	1296	TGCAAGATTCAGTGGTCTACCTGATCTATATATAAAGTAATGAGATAAATAAGAA-	1354
DB	109	LysSerIleHisIleLysLeuAsnAsnLeuPheLeuLysLysAsnGluPheGluLysThr	128
QY	1355	GTAGAAGCCTCCTAAAGCCATCGCTTCAAGCCGCATT-----	1396
DB	129	IleAsnAsnLysIleLysGluIleGluLeuGlnPheLeuIleGlnAsnValThrHis	148
QY	1397	GAAGCAAAACCTCTTCCA-----AATAAGCCTTTGAATGAAGAAATGAACAACTTGA--	1451
DB	149	GluAsnLysGlyLe-ProIleSerLysGluLeuLysGluGlnGluLysHisLeuGlnAs	168
QY	1452	-----GCAGATCCCTGTTCCACCGCAATCCAAACAAAGGACTATAAGAA	1495
DB	168	NserHisGluAsnValAlaIleTyrAspThrHisGluIleGluAsnAspLysLysLy	188
QY	1496	AATTCTTGGGATCTAGAGTCTCTCGAGACTGTTTACAGAGAGGATGTTGTTTACCAAGC	1555
DB	188	SueTyrThrAsnPheHisAsnGluGluLysAspHisLeuLysCysLeuLeuGluGluTy	208
QY	1556	TACACATCAAAAAGATAAGATAAAATAAGAAATAGAAAGTCCCTTAATAAAGTGGTC	1615
DB	208	rSerLysThrLeuGluIleTyrLysMetGlyLysIleGln-----	221
QY	1616	TTCTGAAGCTACCTCGGATGAATTTCTATTCCACTAAAGCCTAGAAATGAAGACATG	1675
DB	222	-----LeuGluPhe-----GluLeuLysCysCy	229
QY	1676	CAAACTTTCAAGCGAGCGCTCCGGGAAGCCATCTGCTTCGAGCCTCCACTGAAATCAAAA	1735
DB	229	SlySGLuLysLeu-----AsnGluGL	236
QY	1736	GTCTGTCCCAATAAGCCTTGAATCAAAAATGAATAACATGAGGAGAGATGAGACTCC	1795
DB	236	uIleGluLysAsnAsnAsnTyrAsnAsnLysMetLysSerTyrGlu-----	251
QY	1796	CATCAGAAATCCAAACAAAGGACTATAAGAAAAATTCGGGATACTGGAGTCTCTGTAGACT	1855
DB	251	-----	251
QY	1856	GTTTCACAGAGANTTGTGTTTACAGAGCTGCGCTCAAAAAGATAGATAAATAATGG	1915
DB		-----	
DB	252	-IleHisIleAspValValLysAsnGluAsnCysLysAsnLeuGluGluLeuAsnAspLe	271
QY	1916	AAAATTAGAGGGTCCCTGTTAAAGTGGTCTCTCAGAGCTAACTGGAATG-----	1966
DB	271	uLysLeuGln-----LeuGluLysThrLysSerGluAsnAsnGlnAsnTyrValLysAs	289
QY	1967	-AAAATTCTATTCCAACTAAAGCCTAGAAATTCATGACATGCAAACTTCAAGACGCGCTC	2025
DB	289	nLysIleLeuAsnAspGluLysAsnLeuAspLysIleAsnAsp-----	305
QY	2026	CCGAGAGCCATCTGCTCGAGCCTCCATTGAAATCAAAAGTCTCTCCAAATACCTTGTG	2085
DB	306	-----LeuLysIleLysIleLysAsnAspLys-----	314
QY	2086	GAATTAAGAATGAACAAACATTCAGGACATGAGAACTCCCATCAAAATCCAAACAAAGGA	2145
DB	315	-----ThrLeuLeuAsnAspAlaGlnAsnLysGluTyrIleLeuAsnAsnPh	330
QY	2146	CTATGAGAAAGTCTTGGGATTCCTGAGTCTCTGTAGACTGTTTCCAGAGATGCTGTT	2205
DB	330	eThrGlnLysIlePheAsnIleIleThrTyrLeuLysAsnAsnAspGluHisAsnPheLe	350
QY	2206	TACCCAGGCTACACATCAAAAGAAATAGATAAAATAATGGAATA-----GAAGAGTC	2259
DB	350	uAsnAspLysIleHisAsnLysLeuAspIleAsnGlnAspGlnIleTyrValGlnThrGl	370
QY	2260	TCTGTAATAGAGTCTTCTGAAGGCTCCCTGAGATGAAAGTTCTATTCCATAAGGCT	2319
DB	370	uLeuTyrIleAspIleIleSerSerIleArgAsnLeuIleAsnPheLysLysThrLe	390
QY	2320	TGAATTCAGGCTCAAACTTTCAAAGCAGCCTCCCGAGAGCCATCTGCTTCCAGGCT	2379
DB	390	uGluArgAsnValGluLeuLysValThrHis-GluMetLysGluLeuArgLysG	410
QY	2380	GCATTGAAATGAAAGTCTGTTCCAAATAAACCCTTGAATTAAGATGAACAAAC-----	2433
DB	410	luleuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	429
QY	2434	---ATTGACACAGATCAGATTTCCCTTCAGATCAAAACAAAGAGGTTGAGAAATCT	2490
DB	430	CysValGluArgAspSerValLysIleAsnSerGluLysGluLysGly-GluLysValI	449
QY	2491	GGGATTCCTCAAGTCTCCGTGGACTGTTTCACAGAAAGGATGTTGTCACCCAGGCTACACA	2550
DB	449	e-----TyrGluLe	452
QY	2551	CAAAAAGAGGATAAATAGTGGAAATTAGAGATTCACAGCTATCAAAATCTTGG	2610
DB	452	uLysGluLysLeuAspAsnAspGluLysIleAsnAspLysLysLysLysLysLysLysLys	470
QY	2611	TACAGTTCATCTGTGAAGGCAAGGAACTTCAAAAAGACACTGTGAACGCTACAGGA	2670
DB	471	-SerTyrGlnValTyrLysMetLysAspTyrGluLysArgGluAsnAsnLeuIleAsnGl	490
QY	2671	AAATGGAACAATGAAAGAGTTTGTGTACTGAAAGAAAGAACTGTC---AGAGAAAAAGA	2727
DB	490	uLysAsnLysLeuLysLeuPheIleGluAsnLysMetThrValGluArgGlyAsnGl	510
QY	2728	AATAATTCAGTACAGAACCAAAAGTTAAATGGGAACAGAGTCTGCGAGTC-----TA	2781
DB	510	uMetAsnAsnLysLysLeuGluGluMetLysGlnLysAsnLysGluLeuIleAsnAsnLe	530
QY	2782	GATTGACTTAAACCAAGAAAGAGAGA-----GAGAAATCCGATATATTAAA-----	2830
DB	530	uAsnAspIleSerAspGluLeuLysAsnCysIleGluGlnValAsnSerValSerArgAs	550
QY	2831	-----GAAAAATTAGAGAAATTAGAGATTCGAGACACCATAGAAAGAG	2877
DB	550	nMetalAsnValGluLysGluAsnIleIleAsn-GluLeuGlnIleLeuArgm	570
QY	2878	TTAGAAGTGAACAACTTGAAGGCTCTCAGATACAGATAAGAAATGAAGGTGTAGA	2937
DB	570	etLysAsnAspThrMetArgLysArgIleSerLysPheValGluGlnGluLysValLeuL	590

QY	2957	-----CTCAACTC	2964
Db	877	eLeuThrAsnHisLeuSerLysGlnLysGluLeuMetAlaGlnLysLysMetAsnSe	897
QY	2965	ATGAAAT	3003
Db	897	rGluAsnSerHisSerHisGluGluGluLysAspLeuSerHisLys-AsnSerMetLeuG	917
QY	3004	AAAGGAATGCCA--TGCAAACTGGAATAGCCACACTA--AACACCAATCCAGAA	3057
Db	917	lnGluGluAlaMetLeuArgLeuGluLeuAspThrLeuLysAsnGlnAsnGlnGluL	937
QY	3058	AAGGAAATAAATCTTTGAGCACTTAAGATTTT--AAAGAAAAGATGCTGAATTTAGAT	3114
Db	937	ysGluLysLysCysPheGluAspLeuLys-ileValLysGluLys-----AsnGluAsp	954
QY	3115	GACCTAAACTGAAAAG--GAATCATTAATAAAGGATCTCA--ATATGT	3162
Db	955	LeuGlnLysThrIleLysGlnAsnGluThrLeuThrLeuThrIleSerGlnTyrAsn	974
QY	3163	GGGCACTAAAGTTCTGATAGCTGAGAAAC--AATGCTCATTTAAATTCAGGAAAAAC	3219
Db	975	GlyArgLeuSerValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGlu	994
QY	3220	AGACAAAGAAA-----CTAGAGCGAATTAATACACATCTGCTGCTGCTGCT	3273
Db	995	LysGlnSerLysGluArgLeuGluAlaGluValGlu--SerTyrHisSerArgLeuAla	1014
QY	3274	CTCAAGCCATCAT-----CAATTGTCATC	3300
Db	1014	laAlaIleHisAspArgAspGlnSerGluThrSerLysArgGluLeuGluLeuAlaPheG	1034
QY	3301	AGAAAAAGTAGAACCTCTCCATTCAGAGAGATGCTTTTGCAAGAAAATGAATG	3360
Db	1034	lnArgAlaArgAspGluCysSerArgLeuGlnAsp-----LysMetAsnP	1049
QY	3361	TTGATGTG--AGATGATGATTAATTAACATGAGTGCTCCATCACCACCTTTCTGAGCTCAA	3417
Db	1049	heAspValSerAsnLeuLys-AspAsnAsnGluLeu-----SerGln	1063
QY	3418	AGCAATCCAAACC--TAAATATTCCTCAATTAGCAGGAGAT-----	3459
Db	1064	GlnLeuPheLysThrGluSerLysLeuAsnSerLeuGluLeuGluPheHisThrArg	1083
QY	3460	-----GTCTAAGAGAATATCATTTGTTTCAGAACATCACAAGAGAC-----	3501
Db	1084	AspAlaLeuArgGluLysThrLeuGlyLeuGluArgValGlnLysAspLeuSerGlnThr	1103
QY	3502	-----AAGCTGAAACACAGTGCAATGAAGGAAGCGAACACATGTATCAAAACACACAGA	3555
Db	1104	GlnCysGlnMetLysGluMetGluGlnLysTyrGlnAsnGluLysVal-----	1121
QY	3556	TAATGGAACAAACACCTGAAACAGCAGATCTCT--AGTCGAAATTAATTTCAACTACAA	3612
Db	1122	-----AsnLysTyrIleGlyLysGlnGluSerValGluArgLeuSerGlnLeuGln	1139
QY	3613	GCAAAA--ATTGTGCTTCACAGCAATTAGTCATGCACATAGAAGCTGACAAACAAAG	3669
Db	1140	SerGluAsnMetLeuLeuArgGlnGlnLeuAspAspAlaHisAsnLysAlaAspAsnLys	1159
QY	3670	CAGATAAATGATATTCATTTCTTGAGAG-----	3699
Db	1160	GluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLysLeuGlnAla	1179
QY	3700	-----GAAATGCAACATCTCTTAAAGAGAAAAGAGGAGGATATT-----AATTAC	3747
Db	1180	GluSerGluLysGlnSerLeuLeuGluGluArgAsnLysGluLeuIleSerGluCys	1199
QY	3748	ATAACCATTTAAAAACCGTATTTCAATATGAAAAGAGAAAGAGAAACAGAACTCATG	3807
Db	1200	AsnHisLeuLysGluArgGlnTyrGlnTyrGluAsnGluLysAlaGluArgGluValVal	1219
QY	3808	AGAACACAGTAGAGAAACTTCTTTGGAGAAACACAGACAGACTTTTACTCACACTCATG	3867
Db	1220	-----ValArgGlnLeuGlnGlnLeuAlaAspThrLeuLysLysGlnSerMet	1236
QY	3868	CTAGAGGCCAGTCTA	3882
Db	1237	SerGluAlaSerLeu	1241
RESULT 7			
ID	Q9HIQ1	PRELIMINARY; PRT; 1710 AA.	
AC	Q9HIQ1;		
DT	01-MAR-2001	(TREMELrel. 16, Created)	
DT	01-MAR-2001	(TREMELrel. 16, Last sequence update)	
DT	01-OCT-2002	(TREMELrel. 22, Last annotation update)	
DE	BA145E8.1	(Hypothetical protein KIAA1074).	
GN	BA145E8.1		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
CC	NCBI_TaxID=9606;		
CX	[1]	SEQUENCE FROM N.A.	
RN	[1]		
RP	Bird C.;		
RA	Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.		
RL	EMBL: AL162272; CAC19649.1; --		
DR	InterPro: IPR002110; ANK.		
DR	Pfam: PF00023; ank; 4.		
DR	SMART: SM00248; ANK; 4.		
DR	PROSITE: PS00688; ANK REPEAT; 4.		
DR	PROSITE: PS02937; ANK REP REGION; 1.		
KW	Hypothetical protein; ANK repeat; Repeat.		
SK	SEQUENCE 1710 AA; 196410 MW; 01CBF9BADB894872 CRC64;		
Alignment Scores:			
Pred. No.:	8,21e-06	Length:	1710
Score:	153.50	Matches:	139
Percent Similarity:	42.16%	Conservative:	84
Best Local Similarity:	26.28%	Mismatches:	165
Query Match:	2.17%	Indels:	142
DB:	4	Gaps:	29
US-09-602-362E-22 (1-4115) x Q9HIQ1 (1-1710)			
QY	2645	AAAGACACTGTGAACACCTACAGGAAATGAACAAATGAAAGAGTTTGTGTACTGA	2704
Db	743	LysAsnHisCysGluLeuLeu-----ThrValIleLysLysMetGluAsp	758
QY	2705	AAAGAACTGTCAAGAACAAAGAAATAATCACAGTAGAGAA-----	2746
Db	759	LysValAsnValLeuGlnArgGluLeuSerGluThrLysGluLysSerGlnLeuGlu	778
QY	2747	---CCAAAAGTTAAATGGGAACAAGAG---TCTGCAGTGTAGATTGACTTAACCAAA---	2797
Db	779	HisGlnLysValGluTrpGluArgGluLeuCysSerLeuArgPheSerLeuAsnGlnGlu	798
QY	2798	GAAGAAGAGAGAGAAATGCCATATATAAGAAATAATTAAGAAATAATAGGAAGATAGAGAAT	2857
Db	799	GluGluLysArgAsnAlaAspThrLeuTyrGluLysIleArgGluGlnLeuArgArg	818
QY	2858	CGAGAGCAGCATAG---AAAGAGTTAGAGTGAAGTGAACACACACTTGAAGGCTCT---	2907
Db	819	LysGluGluGlnTyrArgLysGluValGluValLysGlnGlnLeuLeuSerLeuGln	838
QY	2908	CAGATACAGATAAGAAATTCAGAGTGTAGAAAGTATTTGATAGGTTT-----	2956
Db	839	ThrLeuGlu-MetGluLeuArgThrValLysSerAsnLeuAsnGlnValGlnGluArg	858
QY	2956	-----	2956
Db	858	gAsnAspAlaGlnArgGlnLeuSerArgGluGlnAsnAlaArgMetLeuGlnAspGlyI	878
QY	2957	-----CTCAACTC	2964

Db 878 eluThrAsnHisLeuSerLysGlnLysGluLeuMetAlaGlnLysLysMetAsnSe 898
QY 2965 ATGAAT-----GAATATTATTCTTACATGAAATTCATGTTGA 3003
Db 898 rGluAsnSerHisSerHisGluGluLysAspLeuSerHisLys-AsnSerMetLeuG 918
QY 3004 AAAAGGAATTGCCA---TGCAAAATCGAATAGCAGCACTA---AACACCAATCCAGAA 3057
Db 918 lngluGluAlaMetLeuArgLeuGluLeuAspThrIleLysAsnGlnAsnGlnGluL 938
QY 3058 AAGCAAAATAAATCTTTGAGGACTTAAGATTTT---AAAGAAAAGATGCTCAACTTAGAT 3114
Db 938 ysGluLysLysCysPheGluAspLeuLys-ileValLysGluLys-----AsnGluAsp 955
QY 3115 GACCTTAAATCTGAAAAG-----GAATCATTTACTAAAGGATCTCA---ATATGT 3162
Db 956 LeuGlnLysThrIleLysGlnAsnGluGluThrLeuThrIleSerGlnTyAsn 975
QY 3163 GGGCAGCTTAAAGTTCTGATAGTCAGAAAC---AATGCTCATTTCTAAATTCAGGAAAAC 3219
Db 976 GlyArgLeuSerValLeuThrAlaGluAsnAlaMetLeuAsnSerLysLeuGluAsnGlu 995
QY 3220 AGACAAAGAAA-----CTAGAGGAGAAATGAATACACCATCTGCTGCTCGCT 3273
Db 996 LysGlnSerLysGluArgLeuGluAlaGluValGlu--SerTy:HisSerArgLeuAlaA 1015
QY 3274 GTACAAGCCATGAT-----CAATTGTGACATC 3300
Db 1015 laAlaIleHisAspArgAspGlnSerGluThrSerLysArgGluLeuGluAlaPheG 1035
QY 3301 AAGAAAAGTAAGAACCTGCTTCCACATTTGAGGAGATGCTGTTTGCAGAAAATGAATG 3360
Db 1035 lnA:GalaArgAspGluCysSerArgLeuGlnAsp-----LysMetAsn 1050
QY 3361 TTGATGTC---AGATGACGATATTAAACATGAGTCTCCATCACCATTCTGAGCTCAA 3417
Db 1050 heAspValSerAsnLysLys-AspAsnAsnGluLleLeu-----SerGln 1064
QY 3418 AGGAAATCCAAAACC---TAAATATTCTCAATTAGCAGGAGAT----- 3459
Db 1065 GlnLeuPheLysThrGluSerLysLeuAsnSerLeuGluLleGluPheHisHisThrArg 1084
QY 3460 -----GTCTAAGAGAAATATGTTTTCAGAAATCACAAGAGAC----- 3501
Db 1085 AspAlaLeuArgGluLysThrLeuGluGluArgValGlnLysAspLeuSerGlnThr 1104
QY 3502 -----AAGGTGAAACACAGTCTCAATGAGGAGCGACACATCTCATCAACGACAGA 3555
Db 1105 GlnCysGlnMetLysGluMetGluGlnLysTy:GlnAsnGluGlnValLysVal----- 1122
QY 3556 TAATGGAAACAAACCTGAAACAGCAGAGTCTCT---AGTCAGAAATATTTCACATACAA 3612
Db 1123 -----AsnLysTy:ileGlyLysGlnGluSerValGluGluArgLeuSerGlnLeuGln 1140
QY 3613 GCAAAA---ATTGTGGCTTCACGACATTTAGTCATGCATAGAAAGCTGACAAACAAAG 3669
Db 1141 SerGluAsnMetLeuLeuA:ggGlnGlnLeuAspAspAlaHisAsnLysAlaAspAsnLys 1160
QY 3670 CAGATACAATGATATTCTTCTTGTAGAG----- 3699
Db 1161 GluLysThrValIleAsnIleGlnAspGlnPheHisAlaIleValGlnLysLeuGlnAla 1180
QY 3700 -----GAAATGCAACATCTCTCTTAAAGAGAAAAGAGAGATATT-----AATTAC 3747
Db 1181 GluSerGluLysGlnSerLeuLeuGluGluGluArgAsnLysGluLeuIleSerGluCys 1200
QY 3748 AATACCATTTTAAAAACCGTATATTCAATATGAAAAGAGAGAAAGAGAAACAGAACTCATG 3807
Db 1201 AsnHisLeuLysGluArgGlnTy:GlnAsnGluLysAlaGluArgGluValVal 1220
QY 3808 AGAACAGCAGTAGAAATCTTTTGGAGAAAACACAGACGAGATTTTACTTCACACTCATG 3867

Db 1221 -----ValArgGlnLeuGlnGlnGlnLeuAlaAspThrLeuLysLysGlnSerMet 1237
QY 3868 CTAGAGGCCAGTCTA 3882
Db 1238 SerGluAlaSerLeu 1242
RESULT 8
Q8ILS9
ID Q8ILS9 PRELIMINARY; PRT; 3026 AA.
AC Q8ILS9;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Hypothetical protein.
GN PF14_0165.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
CX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=22255705; PubMed=12368864;
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA Bisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA Fraser C.M., Barrall B.;
RT "Genome sequence of the human malaria parasite Plasmodium
RT falciparum";
RL Nature 419:498-511(2002).
DR EMBL; AB014818; AAN36777.1; -.
KW Hypothetical protein.
SQ SEQUENCE 3026 AA; 357633 MW; 9ECABD915C3C25CE CRC64;
Alignment Scores:
Pred. No.: 2,018-05 Length: 3026
Score: 150.00 Matches: 255
Percent Similarity: 33.77% Conservative: 182
Best Local Similarity: 19.71% Mismatches: 433
Query Match: 2.12% Indels: 424
DB: 5 Gaps: 59
US-09-602-362E-22 (1-4115) x Q8ILS9 (1-3026)
QY 835 CTGATGAGCTGCATCTTGGTGGAGGACATCTGACAAATTCATGTTGGAGAAAGGACA 894
Db 1126 LeuArgLysLeuLysAsnAsnGlyAsnLysCysThrTy:GluLysGlyGluLys 1145
QY 895 TCTGGAAAGTTTCAACGTCAGCAGAGAAAACACCTAGGAAATTCAGTCTGCAAAAAC 954
Db 1146 AsnLysAsnPheCysSerGlnSerArgAsnAsnSerLeuThrSerPheArgAsnLysAsn 1165
QY 955 ATCTGAGAAAT----- 966
Db 1166 AspLysAsnIleLysLysAsnValLeuThrLysLysAspSerIleLeuAsnGlnAspLys 1185
QY 967 TACGGCCAGCAAGAGACCTAGACATCGCAGGAGAAAGACACACACCTAGG 1026
Db 1186 IleAsnAspCysLysAspAspLysLysLysMetGluLysLysLysIleTy:MetAsn 1204
QY 1027 -----AAATATGATCCCGAAAAGAACACTGAGAAATTC----- 1062
Db 1205 SerGlnMetAsnGlnLysLysIleGluSerAspLysLysLysAsnLysPheGluLys 1224
QY 1063 GTGGCCAGAAAGAGACCTAGGAGACGATGGAGA-----AAAGAAACCTGTAAAG 1119
Db 1225 Ty:GlyLysTy:GlyLysAsnGluLysTy:ArgGlyLysTy:HisLysTy:ProLeuAsp 1244
QY 1120 ATGGATGCG-----TGCAAGAGTAACTCTAATAAATAAGTTTTCGAAAAGGAA 1170

1245 ileGluAlaSerTyrA-gCysLysAsn-LeuTyrLeuLysile-AsnValAspThrAsnA 1264
 1171 ATCTAAGATGTTGCATGTCACAAAGAATCATCTACAAAGCAAGTCCCATGATCAGAGG 1230
 1264 sn-----GluGlnIleLeuLysTyrile----- 1271
 1231 TCCCATCAGATCCAAACAAAGGAAGATGAAGAATATCTTTGATTCCTGGGTCTCTTGA 1290
 1272 -----ProAsnLys----- 1281
 1291 AGTTCTGCAAGATTCAAGTGTATACCTGAGTCTATATATAAAGTAATACAGATAATA 1350
 1281 luPheLysLeuIleIleThrLysPheIleAspGlyPheIleLysCysThrGlu----- 1298
 1351 AGAAGTAGAAGCCTCCTAAAAGCCATCTGCCTTCAAGCCGCCATTGAAAGCAAACTCT 1410
 1299 -----SerLysileT 1302
 1411 TTCCAAATAAGCCTTTGAATGAAGAATGAACAAACATGAGCAGATCCGTGTCTCCACC 1470
 1302 yr-----LysSerAsnIleIleLeuGlnAsnIleGluThrAsp---IleSerTyrA 1319
 1471 GAATCCAAACAAAGACTATAGAATAATCTTGGGATCTAGACTCTCTGAGACTGTTT 1530
 1319 snIleThrIleHisAlaTyrAsnThrValValProSerLeu-----TrpLeuT 1335
 1531 ACAGAAGGATTGTGTTTACCAAGGCTACATCAATCAAAAGAAATAGATATAAAATAATGGA 1590
 1335 yrAlaSerIleSerPheTyr-----LeuThrLysPheHisL 1347
 1591 TAGAAGACTCCTTAATAAGTGGTCTTCTGAGGCTACCTCGGATCGAATGAATTTCTATTCC 1650
 1347 euGlu----- 1348
 1651 ACTAAAGCTAGAAATGAAGACA---TGC-----AACTTTCAAGCGAGCCTCCGGGA 1701
 1349 -----AsnPheLysLeuLysThrProCysHisPheThrLeuAlaGlnAsnThrAspPheL 1367
 1702 AGCCATCTGCTTCGAGCCTCCACTGAATCAAAAGTCTGTCCCAATAAGCCTTGGATG 1761
 1367 euAsnSerIleThrAsnAsnAspAsnGluLysLysIleValProLeuLys-----AsnG 1385
 1762 AAAATGAACATCGAGGCGAGATGAGACTCCCATCAGAATCAACAAAGACTATA 1821
 1385 luLysAsnIleThr----- 1390
 1822 AGAAAATTTCTGGGATACTGGAGTCTCTGTAGACTGTTTCACAGAGGATTCGTTTACCA 1881
 1390 InLysile-----IlePheHisAsn-----IleAsnGluL 1400
 1882 AGGTCGCTCAAAAAGATAGATAAAATAATGAAGATTAAGAGGTCCTCTGTATAAGT 1941
 1400 yLysLysMetLysAsnAsnGluLeuThrAsnAsnCysThr-----AsnMetLeuLysA 1418
 1942 GGTCTTCTGAGCTAACTGGGAAGTCTTATTCACAACTTAAAGCCTGAGATGATGA 2001
 1418 spIleGluGluHisAspAsnGln--HisPheTyrLysAsnAsnSer-----T 1433
 2002 CATGMAACTTCAAGACGGCTCCGAGAGCCATCTGCCT---TCGAGCCTCCAT--- 2054
 1433 hrCysSerAspGluAspAsnSerAspGluSerIleValProLysAsnLeuHisIleG 1453
 2055 -----TCAAAATCAAAAGTCTGTCCAAATAAACTTGGAA 2088
 1453 InAsnAsnLeuGluPheLysGlyIleLysAsnGlnLysAsnGlyAsnIleAsnIleSert 1473
 2089 TTAAGATGAACAAACTTGAAG----- 2111
 1473 hrArgAsnAsnValThrLeuAsnAsnAsnIleAsnAsnLeuAsnAsnLysTyrAsnAsnT 1493
 2112 -----CAGATGAGAACT----- 2124

1493 hrIleAspGlyLysThrGluMetAsnSerGlnIleHisAsnThrIleHisGluAspGluG 1513
 2125 CCATCAAAATCCAAACAAAGGACTATGAGAAAGTCTTCGGGATTCCTGGAGTCTCTGTAGAC 2184
 1513 luIleLysSer--ThrSerSerTyrGluGluLeuLeuLeuSerPheHisAspLeuAsnLy 1532
 2185 TGTTTCCAGAGGATGGTGTTCACCCAGGCTACATCAAAAA----- 2227
 1532 sSerGluIleArgAsnLysPheThrGlnValLysTyrGluAsnCysAsnAsnValAspGl 1552
 2227 ----- 2227
 1552 uArgLysileThrProThrPheGluThrProSerGlyIleGluTyrProSerPheCysLy 1572
 2228 -----GA 2229
 1572 sTyrAspTyrAsnSerThrSerAsnGlnLysIleGluValIleAsnLysValCysMetGl 1592
 2230 ATAGATAAAATAATGGAATAATAGAGAGTCTCTGATAATGAGGTTTCTCGAAGGCTCCCT 2289
 1592 uGluValLysAspLysLysLysGluAsnTyrIleValTyrLysLysAsnLysAspAsnLy 1612
 2290 GAGAAATGAAA-----GTTCTATTCCACTPAAGCCTTGAATTCATGCGCATGCAACT 2340
 1612 sGluLeuLysAsnThrThrIleLeuAsnHisAsnProArgAsnLeuAsn-----LysGl 1630
 2341 TTCAAAGCAGGCTCCCGAGAGCCATCTCTTCGAGCCTGCATTCGAAATGAAAGTCTGT 2400
 1630 nSerPheLysPheSer-GluPheSerLeuLysLysLysValAsnLeuAsnThrLysT 1650
 2401 TCCAAATAAACCTTGGAAATTAAGAATGAACAACTTGAACATGAGACGATCAGATTCCCTTCAG 2460
 1650 hrGlnLysAsnIleAsnLysAsnAsp-Asn----- 1660
 2461 AATCAAAACAAAGAGGTTGAGAAATTCGGGATTCCTGAACTCTCCGTGAGCTGTTTCA 2520
 1661 AsnAsnAsnAsnAsnAsnAsnGluAsnAsnAspIle----- 1672
 2521 CAGAAGATGGTGTGTACCCAGGCTACACACAAAGAAAGGATAAAATAGTGGAAATTT 2580
 1673 ---AsnAspIleAspIleSerAsnMetSerSerAsnGlu-----LeuThrile 1687
 2581 AGAAGTACCTAGCTATCAAAATCTCGTACACTTCATCTGTGAAAGGCAAGGAGAC 2640
 1688 LysGluIleThrAspSerAsnSerPheIleLeuSerSerCysLeuSerLysGluAsn 1707
 2641 TTCAAAAGACACTGTGAACACGCTACAGGAAATGCAACATGAAAGAGATTTTGTGTA 2700
 1708 PheLysAsnAsp-----AsnLysLysLysAsnIleGluLeu 1719
 2701 CTGAAAAGAAC-----TGTCAGAACAAAGAAAT 2730
 1720 IleLysSerAsnLysAspSerThrGluAsnSerHisGlnCysGluGlu-LysAsnI 1739
 2731 AATACACAGTACAGAACCAAAAGTAAATGCGAACAAAGACTCTGCAGCTGATGACTT 2790
 1739 eAsnAsnIleGlnGluLys-----GlnThrLeuArgAspIleLeu 1754
 2791 AAACCAAGAAAGAGAGAGA-----GAGAAATCCCGATATATTAAAGA 2832
 1754 uAsnGlnLeuLysGluLysAsnGlyThrTyrAspAspAspLysGlyTyrIleGlnLy 1774
 2833 AAAAATTAGCAAGAAAT----- 2848
 1774 sLysIleHisLysAsnAspAsnIleAsnValAsnCysAsnIleAsnLysTyrAspIleAs 1794
 2849 ----TAGAAGATTCGAGAGCAGCATAGAAAGATTTAGAGTTAGAAACACAACTTGAAGGC 2904
 1794 pLysAsnLysAsnIleSerIleAsnAsnAsnThrIleAsnAsnAsnThr-----I 1811
 2905 TCTCAGATACAGATGAAGATTCAGAGGTGATGAAGTAAATTTGAATAGTTCTTCAACTC 2964
 1811 eAsnAsnAsnThrValAsnAsnLysTyrIleLysLysGluHisAsnSerPheIleAsnTy 1831


```
Db 115 heThrSerSerPheIleLeuPheHisSerPheThrSerSerPheIleLeuPheHisSerP 135
QY 3456 TCCTGCTAATGAGATAATTTTAGG---TTTGGATTCTCTTTGAGCTGAGAAAGTGGT 3400
Db 135 heThrSerSerPheIleLeuPheHisSerPheThrSerPheIleLeu-----151
QY 3399 GATGGAGCACTCATTTGTTAATATCGTACTCTCAGATCAACATTCATTTCTTCCAAACA 3340
Db 152 -----PheHisSerPheThrSerSerPheIleLeuPheHisSerP 165
QY 3339 GCATCTCTCAATGTGGAAGCAGGTCTTACTTTTCTGTATGTCACAAATTGATCATGGC 3280
Db 165 heThrSerSer-----168
QY 3279 TTGTACAGCGAAGCCAGTCAGATGGTGTATTCAATTTCTCCCTCTAGTTTCTTTCTCT 3220
Db 169 -----PheIleL 171
QY 3219 GTTTTCTCAATTAGAAATGAGCAATGTTTCTCAGCTATCAGAACTTTTAGCTGCCACACA 3160
Db 171 euPhe-----HisSerPheThrSerSerPheIleLeuPheHisSerP 185
QY 3159 TATTGAGATCCCTTTTAGTAATGATTCCTTTTCAGTTTATAGGTCATCTAAGTTTCAGCAT 3100
Db 185 heThrSerSerPheIleLeuPheHisSerPheThrPheIleIleHisPheIleSerPheP 205
QY 3099 CTTTCTCTTTAAATCTTAAGTCCTCAAGATTTATTTCTCTTTCTCGGATTCGTGTTA 3040
Db 205 heTyVal-----HisSerTyTyPhe-----IleLeuLeuA 217
QY 3039 GTGTGGCTATTCAGTTTTCATGGCAATTCCTTTTCAACATGCAATTTTCATGTAAGAA 2980
Db 217 rgLeuSerPheIleLeuPheHis-----PhePheThrSerSerPhe-----I 231
QY 2979 TAATTTTCAATTCATGATGAGTTCAGAAACCTATTCAATTAATTTCTACCTTCATCTT 2920
Db 231 leuPheHisSer-----LeuThrLeuSerPheIleLeuPheHisSerL 246
QY 2919 ATCTTGATCTGAGAGCCTTCAAGTTGTCTTCACTTCTAATCTTTCTATCTCTCTC 2861
Db 246 euThrLeuSerPheIleLeu-----PheHisSerLeuThrLeuSerPheIleLeuP 263
QY 2860 -----TCGATCTCTCAATTCCTTCAATTTTCTTAAATATATATATCGG 2818
Db 263 heHisSerPheThrPheIleIleHisIleIleSerPhePheTyTyArgLeuPheThr 283
QY 2817 CATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2758
Db 283 yrPheIleSer-----LeuLeuProA 290
QY 2757 TTAACCTTTGGTCTCTACTGTGATTTATTTCTTTTGTCTGCACAGTCTCTTTTTCAGTAC 2698
Db 290 rgSerPheTyTyPhe-IleLeuLeuArgSerSerPheIleLeu-----PheHisSerPhe 307
QY 2697 ACAAACACTCTTTTTCATGCTTCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2647
Db 308 ThrPheIleIleHisIleSerPhePheTyTyValCys-HisSerLeuAsnPheIleLe 327
QY 2646 TTGAAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2587
Db 327 uLeuArgLeuSerPheIle-----LeuPheHisSerPheThrPheIleValHi 343
QY 2586 TCTTCTAAATTTCCACTATTATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2529
Db 343 sPheIleSerPheTyTyPheTyArgSerTyTyPhe-----PheThrPro 358
```

RESULT 11

```
Q86KF8
ID Q86KF8 PRELIMINARY; PRT; 1781 AA.
AC Q86KF8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
```

```
DE Similar to Dictyostelium discoideum (Slime mold). Interaprin.
OS Dictyostelium discoideum (Slime mold)
OC Eukaryota; Eukaryota; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum."
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR ENBL; AC115680; AAC51092.1; -.
SQ SEQUENCE 1781 AA; 209555 MW; 6483BE19ED1B76A8 CRC64;
```

Alignment Scores:

Pred. No.:	7.25e-05	Length:	1781
Score:	144.00	Matches:	237
Percent Similarity:	38.21%	Conservative:	202
Best Local Similarity:	20.63%	Mismatches:	395
Query Match:	2.03%	Indels:	316
DB:	5	Gaps:	49

US-09-602-362E-22 (1-4115) x Q86KF8 (1-1781)

```
QY 817 AGCTGCTGGGAAAACACCTGATGAGCTGCATCTTGGT-----GGA 858
Db 272 LysLeuLysGluSerGluIleLeuMetAspLeuGlnIleGlnValSerAspLeuGln 291
QY 859 GAAACATCTGACAAATTCATTTGCGAAGAGGACATCTCGAAAGTTCGACGTCAGAG 918
Db 292 SerThrSerAspAspAsn-----GluArgTyGlnSerLeuIleSerGluTyGln 309
QY 919 AAAACACCTAGGAAATTCGATCTGCAAAAACATCTGAGAAATTTACGGCCAGCA 978
Db 310 GluAsnLeuLysSerValSerGlnLeuAsnLysGluIle-----322
QY 979 AAGGAAGACCTAGAGATCGCAGAGAGAGAGACACACCTAGGAAATATGAGTCC 1038
Db 323 AsnGluLysLeuLysAsnGluArgThrGluSerLysSerLeuIleAspLys-IleSerLy 342
QY 1039 CGAAAAGAAACACTGAGAAATTCGTCGGCAGAAAAGGAA-----GACCTAGG 1086
Db 342 sGluAsnAsnAspTyTyGlnIleLeuIleThrGluGlnAspLysAspLeuLys 362
QY 1087 AAGACGCATGGGAGAAA-----GAAACCTGTAAAGATGATGCGTGC 1131
Db 362 sSerGlnLeuAspThrLysSerAsnAsnTyTySerLysLeuLeuAspLeuSerLeuSe 382
QY 1132 AAGAGTAACTCTATAAATACTTAAAGTTTGAAGAAAGAAATCTAAGATGTTGTCATGCCA 1191
Db 382 rLysLeuAsnTyTyAsnLysLeu-AsnLeuGluPheAsnAsnIle-----396
QY 1192 CAAAAGATCATCTACAAAAGACGATGCCATGATCAGAGTCCCATCAGATCCAAACAAGG 1251
Db 397 -----SerAsnGluTyTyGlnIleGluLysGlnLeuSerPheAspPheAsnGluL 414
QY 1252 GAAGATCAAGAATATCTTTGATTTCTCGGCTCTTTTGAAGTCTTCTGCAAGATTCAGTGG 1311
Db 414 euLysLysLysLeu-----AspAspGlnIleGluArgAsnSerAsnGlnSerLeu- 431
QY 1312 TATACCTGAGTCTATATATAAAGTAATGAGATAAATAAGAGTAAAGACCTCTCTATAA 1371
Db 432 -----LeuThrGlnArgIleAsnGluIleIleAspLysGlnLysGluIleAspA 448
QY 1372 AGCCATCTGCTTCAAGCCGCCCATTTGAACCAAACTCTTTTCCAAATATAGCCCTTGAATG 1431
Db 1431
```

Db 448 snLeuAsnSerAspLeuGlnSerLeuAsnAspLys-----GlnSerAspSerAsnA 465
QY 1432 AAGAATGAACAACATGTAAGCAGATCCGTGTCCACCGAATCCAAACAAGAC----- 1487
Db 465 spGlnIleAsnThrIleThrAsnGluAsnAspLeuLysIleLeuGluAsnSerI 485
QY 1488 -----TATAAGAAAATTCCTGGGATTCAGAGACTCTGAGACTG 1527
Db 485 leLeuLeuIleSerAsnHisGlnAspLysLeuThrThrAsnLysLysGluLeuGluSerL 505
QY 1528 TTTACAGAGGATGTTTACCAAGCTACACATCAA-----AAAG 1569
Db 505 euTyrSerIleIleGluAsnLeuAsnSerThrHisGlnAspGlnLeuLysGluAs 525
QY 1570 AATAGATAAAAAATGCAAAATAGAA-----GAGTCCCTTAATAAGTGGTCTTCTGAAG 1623
Db 525 snGlnIleGluGlnMetLysLeuAspGlnSerGluSerLeuAsnLysPheGlnGlu-Lys 544
QY 1624 GCTACCTCGAATCAAAATTCATCCACTAAAGCCTAGAAATGAGACATGCAAACTTT 1683
Db 545 ileThrThrLeu-----AsnHisThrIleValAsnIle 555
QY 1684 CAAAGCGAGCTCCGGAAGCCATCTGCTT-----CGAGCTCCACTGAAATCAAAAGT 1737
Db 556 AsnSerGluLysAspSerLeuAsnCysLeuIleAsnAspTyrSerGluGlnIleAsnGlu 575
QY 1738 CTGTCCCAATAGCCTTGGATGAAATGAAATGAAACATGAGGAGCAGTGAAGACTCCA 1797
Db 576 LeuAsnLysIleAsn-----AsnGlnLysAsn-Gln----- 585
QY 1798 TCAGAAATCCAAACAAAGGACTATAAGAAAATTCGGGATCTGAGTCTCTGTAGACTGT 1857
Db 586 -----GluAsnGlnGlnLeuLeuSerSerIle----- 594
QY 1858 TTCACAGAGGATGTTGTTTACCAAGGCTCGCTCAAAAGAAATAGATAAAATATGGAA 1917
Db 595 -----ArgGluAspCys----- 598
QY 1918 AATTAGAGGCTCCCTGTTTAAAGTGTCTTCTGAGGCTAACTGGGAATGAAAGTCTATT 1977
Db 599 -----GlnThrIleAs 602
QY 1978 CCAACTAAAGCCTAGAATGATGATCATGCNA-----ACTTCRAAGCAGGCTCCCGAGAGCC 2034
Db 602 nGlnLeuLysSerThrIleLysSerLeuGlnAspGluSerAsnSerLeuSerThrGluI 622
QY 2035 ATCTGCTTCGAG-----CCTCCATTCGAAATCAAAAGTCTGTCCAAATAAACCTTG 2085
Db 622 eGluAlaIleLysLeuGlnLeuAsnGlnLeuSerThrIleThrIleProGluLys----- 640
QY 2086 GAATTAAGATGAACAACATTTGAGGAGATGAGACACTCCCATCAATCCAAACAAGGA 2145
Db 641 -AspGlnGluLeuSerAsnLysGluArgThrIleGlnGluPheGlnValLysThrGlnG 660
QY 2146 CTATGAGAAGTT-----CTTGGGATTCGAGTCTCTGTAGACTGTTTC 2190
Db 660 nLeuLysGlnThrIleGlnGlnAsnGlnLeuThrIleAsnGlnHisLeuThrThrIleAs 680
QY 2191 CAGAAGGATGTTTACCCAGGCTACACATCAAAAGAAATAGATAAAATATGGAAT 2250
Db 680 pAsnGlnSerValAspIleAsnSerLeuAsnGluLys----- 692
QY 2251 AGAAGAGTCTCTGATAATGAGTCTTCTCAAGGCTCCCTGAGATGAAAGTCTTATTCCA 2310
Db 693 -----LeuValGlnLeuAsnAspGluSerIleLysLysGlnGlnSerIleHi 708
QY 2311 CTAAGCCTTGATTCATGATGCGATGCAAACTTCAAGCAGCGCTCC----- 2356
Db 708 s-----SerLeuSerLeuGlnValIleGluLeuAsnLysLysLeuSerGluLysAspG 727
QY 2356 ----- 2356
Db 727 nTyrAsnGlnSerLeuGluSerIleAspGlnLeuThrSerGluLeuGlnLeuLysGlnAs 747

QY 2357 -----CGAGAGCCATCTGCTTCGAGCCTCGATTGAAATGAAAAGTCTGT 2400
Db 747 pAspLeuAsnArgGlnGlnGlu-GlnLeuGlnLysAsnSerIleAspIleAspGlnLeuP 767
QY 2401 TC---CAAATAACTTTGGAATTAAGAAATGAACAACATGAGACAGATGATTTCCCTT 2457
Db 767 heAspLysIleAsnLeuGlyLysSerAsnPheAspGluLeuAsnGlnArgTyrGlnValG 787
QY 2458 CAGATCAAAACAAAGAGGTTGAGAAATTCGGGATTCCTGAGTCTCGTGGAGTGT 2517
Db 787 luginAsnGlnLeuPheAsnLeuLysGln-----AspLeuGlnGlnSerIleAsnLeuP 805
QY 2518 TCACAGAGGATGTTGTATCCAGGCTACACACAAAAGAAAGGATAAAATAGTGGAAA 2577
Db 805 heAsnGluSer-----LysLeuTyrThrThrGlnLeuGluLysSerIleGluG 821
QY 2578 ATTAGAAG-----ATTCACTAGCCTATCAAAATCTTCGTATC 2613
Db 821 InValLysArgValAsnGluGlnSerAsnHisGlnGlyGluLeuThrIleLeuLeuA 841
QY 2614 AGTTCATCTTGTGAAAGCAAGGAACTT----- 2642
Db 841 spGluGluArgTyrLysCysGlnThrLeuLysMetAspLeuGluAsnSerLysGlnIleA 861
QY 2643 -----CAAAAAGACACTGTGAACACGTAC---AGGAAAATGGAACAATGAAAAGAG 2691
Db 861 snGlnAsnGlnSerAspAspLeuAsnPheTyrAsnAsnGlnIleGluGlnLeuLysGlnG 881
QY 2692 TTTTGTGTACTGAAAAAAGAACTGTCAAGACAAAAGAAA----- 2729
Db 881 InLeuSerGluSerLysGluLeuValGlnGluCysLysLeuSerIleSerGlnPheGluL 901
QY 2730 -----TAAATCACAGTAGAGAACCAAAAGTTAAATGGAACAAG----- 2768
Db 901 ysThrValIleGluGlnLeuGluIleLysGluLeuSerAsnGluLysGlnGlnTyrG 921
QY 2769 -----AGTCTGCGTGTAGATTGAGTCTTAAACCAAGAAAAGAGAGAGAA 2814
Db 921 InGluThrCysGlnSerLeuSerLeuLysLeu-----SerLysLeuAsnAspGlnSerA 939
QY 2815 ATCCGATATATTAAAGAAAATTAAGAAATTAAGAAATCGAGAG-----AGCA 2868
Db 939 sn-----AspGlnLeuGluGlnIleGlnGlnLeuGlnSerAsnSerL 954
QY 2869 TAGAAAGATTAGAAAGTGAACACAACTTGAAGGCTCTCAGATACAGATAGAATT-G 2927
Db 954 euAspLeuGlnAsnSerGlnAsnGlnIleSerLeuLeuGlnAspSerLeuAsnGluThrS 974
QY 2928 AAGGTGTAAGAAATTAATTAAGTCTTCAACTCATGAAATGAAATTTATCTTACA 2987
Db 974 eAspLeuLysSerGlnLeuGlnCysGlnLeuAsnGluSerAsnGluIleIleSerLysL 994
QY 2988 TGAATTCATGTTTGAAGAAAGTATCCCATGCAAACTGGAATAGCCACCTAAACACC 3047
Db 994 euGluLeuLysIleGlu-----GluAsnGlnAsnGlnLeuThrGluPheG 1009
QY 3048 AATCCAGAAAAGAAAATAAATCTTTTGAAGGACTTAAAGATTTTAAAGAAAAGATGCTGA 3107
Db 1009 luLeuLysIleGlnSerLeuSer----- 1016
QY 3108 CTTAGATGACCTTAAACTGAAAGGAATCATTTACTTAAAGGATCTC-----AATATG 3161
Db 1017 -----SerGlnTyr-AsnGlnAspLeuGlnGlnIle 1027
QY 3162 TGGCAGCTAAAGTCTGTAGCTGAGAAAACAATGCTCATTTCTAAATTCAGAAAAAAGAG 3221
Db 1028 GluGlnSerLysMetLeuIleAspGluLysGlnSerCysIleGlnLeuGlnMetGlu 1047
QY 3222 ACAAGAAAACCTAGAGGACAGAAATTAATATACACCATCTCGACTCGGTCTGCTGTACAGC 3281
Db 1048 IleAspLysAsnAsn-----HisLysIleGlnGlnLeuGln----- 1059

```

QY 3282 CATGATCAATTGTGACATCAGAAAAAGTAAGAACCTGCTTCCACATTGAGGAGATGCTG 3341
Db 1060 -----GlnAspLeuSerThrSerAspTyrLysIleGlnGlnLeuGln 1073
QY 3342 TTTCAGAGAAATGAATGTTGATGTGAGAGTACGATATTAACAATCAGTGTCTCCATCAC 3401
Db 1074 IleAspLeuGlnIleAspLysAspGluIleIleLysLeuGluGluThrlle----- 1090
QY 3402 CACTTTCTGAGCTCAAGGAAATCCAAAACCTTAAATTAATCTCAATTACAGAGATGT 3461
Db 1091 -----SerGlnArgAsnGlnSerIleLys-----GluSer 1100
QY 3462 CTRAGAGAAATACATTTGTTTCAACATCACAAAGAGACAAACGCTGAAAACAGTCAAT 3521
Db 1101 LeuValLysCysAsnAspLeuGlnAspGluThrSerLysLeuAsnAspAsnLeuGln 1120
QY 3522 GAAGGAGCAACACATGTTATCAAAACGACACAGATATGGAACAAACACCTGCAACAGCAG 3581
Db 1121 LeuAsnSerThrIleThrAspTyrGlnSerGlnIleThrGluSerAsnGluAsnValGln 1140
QY 3582 AGTCTCTAGTCAGAAATATTTCAACTACAAAGCAAAAATTTGTGCTTCACAGCAATTAGT 3641
Db 1141 SerLeuGlnAsnGlu-LysAsnGlnLeuGlnLeuGluLeu-----AspGlnLeuLys 1157
QY 3642 CATGCAATAGAAGCTGACACAAAGAGCAGATACACATGATATTCATTTCTTGAGAGGA 3701
Db 1157 sGlnArgIle---SerGluGlnHisAspAspIleThrLeu----- 1169
QY 3702 AATGCAACATCTCTCTTAAAGAGAAAGAGAGAGATATTAATTAACATACCAATTAAAA 3761
Db 1170 -LeuAsnSerIleGluPheGluLeuAsnArgLysIleSerAsnTyrGlnSerAspIleLys 1189
QY 3762 ACCGTAT-----ATTCAATATGAAAAGAGAAA-----GAGAA 3794
Db 1189 sGluTyrAspAsnAsnIleLysValIleGlnAsnGluLysAsnGlnLeuGluLeu 1209
QY 3795 ACAGAACTCATGACACAGACGATAGAACTCTTTGGA---GAAACACAGACAGAC 3851
Db 1209 uAspGlnLeuLysGlnValLeuSerAspLysGlnAspGlyValSerThrLeuAsnSerTh 1229
QY 3852 TTTTACTCACACTC 3864
Db 1229 rLeuLeuGluLeu 1233

```

RESULT 12

```

Q81BY8      PRELIMINARY;      PRT;      2910 AA.
AC Q81BY8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN PF07_0042.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]_TaxID=36329;
RP SEQUENCE FROM N.A.
RA Seeger K., Murphy L., Harris D., Bertiman M., Fain A., Hall N.,
RA Quail M., Barrell B.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL844506; CAD50853.1; -
KW Hypothetical protein.
SQ SEQUENCE 2910 AA; 350698 MW; B9F142C5A895C8C1 CRC64;

```

Alignment Scores:

```

Pred. No.:      7,86e-05      Length:      2910
Score:          144.00      Matches:      223
Percent Similarity: 38.07%      Conservative: 200
Best Local Similarity: 20.07%      Mismatches: 357
Query Match:      2.03%      Indels:      331
DB:              5          Gaps:      51

```

```

US-09-602-362E-22 (1-4115) x Q81BY8 (1-2910)
QY 1133 AGAGTAACATCTAATAAATAAGTTTGAATAAGGAAATCTAAGATGTTGATGTCATCCAC 1192
Db 1072 LysIleLysTyrAsnAspIleLysThrAsnLeuGluA:ggLulysLysIleasnMethis 1091
QY 1193 AAAAGAATCATCTACAAAGCAAGTCCCATGATCAGAGGTCCCATCAGATCAGATCCAAACAGG 1252
Db 1092 MetAsnIleGluLeuGluLysGluArgLysIleAsnAspGluMetMetIleLysPheGlu 1111
QY 1253 AAGATGAAGAATATTCTTGTATTCCTGGGTCTCTTTGAAGTCTCTCAAGATCAAGTGT 1312
Db 1112 Lys---LysAsnLysMet---TyrAspAspIleAsnLysGluLeuGluAsnLysLys 1129
QY 1313 ATACCTGAGTCTPATATATAAAAGTAATGAGATAAATAAGAGTAGAAGAGCTCTCTATAAA 1372
Db 1130 GlnAsnLysGluIleTyrMet-----GluLeuGluAsn---IleLys 1142
QY 1373 GCATC-----TGCCTTCAAGCCCATTCGAAGCAAAACTCTTTCCAATAAGCC--- 1423
Db 1143 AsnIleAsnSerAspMetAsnPheAlaLeuGluAspLysSerValGlnTyrAspGlnLys 1162
QY 1424 -----TTTGAATGAAGAATGAACAAACATTCGAAGCAGATCCGTGTCTCC----- 1466
Db 1163 CysIleAspPheGlu---LysLeuAsnLys-MetTyrGluAspThrCysLysGluLeuAs 1181
QY 1467 -----CACGAATCCAAACAAAGGACTATTAAG----- 1493
Db 1181 nAsnLysAsnValLeuTyrGluGlnMetAsnThrGlnLeuTyrLysGluLysIleAsnAs 1201
QY 1494 ---AAAATTCTTGGGATTTCTGAGATCTCTCGAGACTGTTTACAGAGGATGTTGTTTAC 1549
Db 1201 nAspGluIleCysThrGluLeuGluLysIleLysAsnIleAsnLysIleGlnIleAs 1221
QY 1550 C-----AAGGCTACATCAATAAAGATAAGATAAATAAATAAGATAAGATAAGATAAG 1597
Db 1221 nLeuGluAsnGluLysThrAsnGluGlnLys-AspLysGlnIleGluLysGlnLys 1240
QY 1598 GTCCCTTAATAAGTGTCTTCTGAAGGCTACTCTCGG-----AATGAATTT 1642
Db 1241 MetAsnLysGluMetValIleGlnLeuGluLysGluLysIleIleAsnAsnGluIle 1260
QY 1643 TCTATTCACATAAAGCCCTAGAAATTGAAGACATGCAAACTTTCAAGGAGGCTCCGGAA 1702
Db 1261 IleThrGlnLeuGlu-----AsnGluLysLeuAsnAsn 1271
QY 1703 GCCATCTGCTTCGAGCCTCCACTGAATCAAAAGTCTGCCAAATAGCTTGAATGA 1762
Db 1272 LysIleIleTyrGluAsp-----IleGluLys 1280
QY 1763 AAAATGAAA---AACATGGAGGAGATCAGAACTCCCATCAGAAATCCAAACAAAGGACTA 1819
Db 1281 LysAsnLysLeuAsnSerGluLeuGluAsn---TyrGluAsnGlnGlnLysLysIle 1299
QY 1820 TAAGAAAATTTCTGGGATCTCTGTAGACTGTTTTCACAGAGGATGTGTGTTT-- 1877
Db 1300 AsnGluMetIleIleGlnLeuGluLysGlu-LysIleIleAsnAsnGluIleIleGln 1319
QY 1878 -----ACCAAGGCTCGCTCAAAAAGATAGATATAAATAATGAAAATTTGAAGG 1927
Db 1319 nLeuGluAsnGluLysGlySerValLysLys-----IleAsnThrGluLeuGlu-- 1335
QY 1928 GTCCCTGTTAAAGTGTCTTCTGAG-----GCTAACTGGGAATGAAAGTTCTA 1975
Db 1336 -AsnIleLysLysMetAsnAspGluMetAsnGluLysLeuAsnLysGluGlnLysIleAs 1355
QY 1976 TTCCAATAAGCTAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2035
Db 1355 nAsnAspLeuGlnLeuAspLeuGlu----- 1363
QY 2036 TCTGCCTTCGAGCCTCCATTGAAATCAAAAGTCTGTCCAATAAATCACTTGA-- 2087
Db 1364 -----AsnGluLysAsnGluLysAspHisIleAsnLysGluPh 1376

```


Db 1999 lylusGln---AlaAsnLysLysLeuAsnIleLeu-----LeuAspGlnAsnLysL 2015
QY 3932 TTTTAGAGAA 3942
Db 2015 ysileAsnGlu 2018
RESULT 13
Q8CIP5 PRELIMINARY; PRT; 1345 AA.
AC Q8CIP5
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Dispatched B.
GN Disp2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=129;
RX MEDLINE=22259598; PubMed=12372301;
RA Ma Y., Erknur A., Gong R., Rao S., Taipale J., Basler K., Beachy P.,
RT "Hedgehog-Mediated Patterning of the Mammalian Embryo Requires
RL Transporter-Like Function of Dispatched.";
RL Cell 111:63-75(2002).
DR EMBL; AY150699; AAN52162.1; -.
DR MGD; MGI:2388733; Disp2.
DR InterPro; IPR000731; SSD_5TM.
DR PROSITE; P50156; SSD; 1.
SQ SEQUENCE 1345 AA; 147943 MW; 931738C26FD3A57F CRC64;

Alignment Scores:

Prod. No.: 0.000173 Length: 1345
Score: 140.00 Matches: 251
Percent Similarity: 31.90% Conservative: 171
Best Local Similarity: 18.97% Mismatches: 414
Query Match: 1.95% Indels: 489
DB: 11 Gaps: 60

US-09-602-362E-22 (1-4115) x Q8CIP5 (1-1345)

QY 3355 ATTTCCTTTGCAACAGCATCTCTCAATGTGGAAGCAGGTCTTACTTTTCTTGATGT 3296
Db 245 ValPheValSerThrSerGlySerLeuTrpAsnLeuGlnAlaIleHisSerMetCys 264
QY 3295 CACAATTGATCATGGCTTCTCAGGAGCAGCCAGTCAGGATGGTGATTCAAATTTCTGCCT 3236
Db 265 Arg-----IleGluGlnGlnIleArgSerHisIleSerPheGlyAla 279
QY 3235 CTAGTTTCTTGTCTGTTTTCTCAATTTAGAATGAGCATTTGT- 3191
Db 280 Leu-----CysGlnArgSerAlaAlaAsnGluCysCysProSerTrpSerLeu 295
QY 3191 ----- 3191
Db 296 GlyAsnTyrLeuAlaValLeuSerAsnArgSerSerCysGlnAspThrThrGlnAlaAsp 315
QY 3190 -----TTCTCAGCTATCAGAACTTTAGCTGCCACATATTGAGATCCCTTTTA 3143
Db 316 ThrAspArgThrLeuAlaLeuLeuArgPheCysAlaThrPheTyrHisArgGlyValLeu 335
QY 3142 GTA-----ATGATTCCT--- 3131
Db 336 ValProAlaCysValGlySerSerGlnAspLysProProPheCysAlaGlnValProAla 355
QY 3130 -----TTTCAGTTTTAGGGTCATCTAAGTTTCAGCATCTTTT 3095
Db 356 LysCysThrGlySerAsnValValTyrGluPheLeuHisTyrLeuLeuAspArgAspPhe 375
QY 3094 CTTTAAATCTTAAGTCTCTCAAGATTTATTTCTCTTCTCGGATTTGTTAGTGTG 3035

Db 376 -----LeuSerProGlnThrAlaAspTyrGlnValProSerLeuLysPheAlaLeu 392
QY 3034 GCT---ATTCCAGATTTTCATGGCAATTCCTTTTCAACATGCATCAATTCATGTAAGAATA 2978
Db 393 LeuPheLeuProIleIleLysThrSerSerLeuLeuAsp----- 405
QY 2977 ATTTTCATTTTCATGAGTTGAGAAACCTATTCAAAATTACT---TTCTACACCTTCAATTCCT 2921
Db 406 IletyrLeuAspGlyLeuGlyAspProIleLysValSerAsnTyrThrSerIleSer 425
QY 2920 -----TATCTTGATCTGAGAGCCTT 2900
Db 426 GlyMetAspLeuGlyLeuLysProArgLeuLeuLysTyrTyrLeuAlaGlnAspThrMet 445
QY 2899 TCAAGTTGTTTTCATCTTCTAACTCTTTCTATGCTCTCTCGATTTCTTCTAATTTCTTCTCT 2840
Db 446 TyrProLeuIleAlaLeuValIle----- 454
QY 2839 AATTTTCTTTTAATATATCGCATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2780
Db 455 -----PhePheGlyMetSerLeuTyrLeuArgSerLeuPhe----- 466
QY 2779 CACTGCAGACTCTTTTCCCAATTTAACTTTTGGTTTCTCTACTGTGATTTATTTCTTTTGT 2720
Db 467 -----Ile-ThrPheMetSerLeuLeuGlyValLeuGlySerLe 479
QY 2719 TCTGACAGTCTTTTTCAGTACACAAAACCTCTTTTTCATTTGTTCCATTTTCTCT---GTAC 2663
Db 479 uMetValAlaTyrPheLeuTyr-HisValAlaPheArgMetAlaTyrPheProPheValA 499
QY 2662 GTGTTTCACAGT---GTCTTTTGAAGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606
Db 499 snLeuAlaAlaLeuLeuLeuLeuSerGlyValCysValAsnTyrThrLeuIlePheLeuA 519
QY 2605 ATTTTGTAGGCTA-----GTGAATC 2585
Db 519 spMetTrpArgLeuSerArgGlyGlnValProSerGlyGlyMetProHisArgValGlyA 539
QY 2584 TTCTAATTTTCCACTATTTCCTCTTTTCTTTGTTGTTAGCTGGGTACACACCATCCT 2525
Db 539 rgThrMetHisPheGlyTyrLeu-LeuLeuValSerGlyLeu----- 553
QY 2524 TCTGTGAACAGTCCACGAGACTTCAGAAATCCAGAAATTTCTCAACCTTCTT--- 2471
Db 554 -----ThrThrSerAlaAlaPheTyrGlySerTyrLeuSerArgLeuProAla 569
QY 2470 -----TGTTTGTATCTGAGGAAATCTGATCTGTCTCAATGTTGTTCTTCTTAATT 2417
Db 570 ValArgCysPhe----- 573
QY 2416 CCAAGTTTATTTTGGAAACAGACTTTTCATTTCAATGCAGGCTCGAAGCAGATGGCTCG 2357
Db 574 AlaLeuPheMetGlyThrAlaValLeuValHisMetGlyLeuThrLeuLeuTrpLeuPro 593
QY 2356 GGAGGCTGCTTTGAAAGTTTGATGCCATGCATCAATCAAGCTTTAGTGAATAGACTTT 2297
Db 594 AlaThrValVal-----LeuHisGlu-----ArgTyr 602
QY 2296 CATTTCTCAGGAGCCTTCAGAAACCTCATTATCAGAGACTCTTCTATTTTTCCATTATTT 2237
Db 603 LeuAlaHisGlyCysValAlaGlnAlaHisGlnArgGlyGlySerAspProLeuArg 622
QY 2236 TATCTATCTTTTGTGATGTAGCTGGGTAACACCATCTCTTCTCGAAACAGTCTACAG 2177
Db 623 LeuLeuAlaLeu-----His-----Arg 629
QY 2176 AGACTCCAGAATCCCAAGAACTTTCTCATAGTCTCTTTGTTGGATTTGATGGAGTTCTC 2117
Db 630 ArgIleArgIlePheArgLysIleIleSerIleLeuSerArgLeuLeuPheGlnArgLeu 649
QY 2116 ATCTGCTCAATGTTTGTTCATTTCTTAATTCCAAGGTTTATTTGGACAGACTTTTGAATT 2057


```

QY 369 TGGCACCACGTACAAAATCCACTATATAACGCATAATGGAAGCC-----AT 325
Db 1278 nGlyAArgspThrLeu---TrpLeuAlaLeuLysGluThrIleTyrAspProAsnMe 1297
QY 324 GTTGATACATACATCAGAGATTTATATCGGCACCGAATCTATCAATATTTGCAAGCCCT 265
Db 1297 tProAsnSerHisHis-----SerSerLe 1305
QY 264 CCTGTGCATGTGAAG-----CCTTCATCAGA 238
Db 1305 uSerTrpLysGlyArgGlyGlyProGlyAspIleSerProValMetLeuProAsnSerGI 1325
QY 237 G-----GTGTCCCTGTTCGCCAAGGACGTGAGCTGGCAGCTCTGCTAC 192
Db 1325 nProAspLeuProAspValTrpLeuArgArgProSerThrTyrThrSerGlyTyr 1343

RESULT 14
Q7YVS2 PRELIMINARY; PRT; 1679 AA.
AC Q7YVS2;
DT 01-OCT-2003 (T-EMBLrel. 25, Created)
DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)
DE 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypochemical protein.
GN TB927.2.1530.
OS Trypanosoma brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5691;
RN [1]
RF SEQUENCE FROM N.A.
RC STRAIN=GUT10.1;
RA El-Sayed N.M.A., Ghedin E., Song J., MacLeod A., Bringaud F.,
RA Larkin C., Wanless D., Peterson J., Hou L., Taylor S., Tweedie A.,
RA Biteau N., Khalak H.G., Lin X., Mason T., Hannick L., Caler E.,
RA Blandin G., Bartholomeu D., Simpson A.J., Kaul S., Zhao H., Bai G.,
RA Van Aken S., Uterback T., Haas B., Koo H.L., Umayam L., Suh B.,
RA Gerrard C., Leech V., Qi R., Zhou S., Schwartz D., Feldblyum T.,
RA Salzberg S., Tait A., Turner M.R., Ullu E., White O., Melville S.,
RA Adams M.D., Fraser C.M., Donelson J.E.;
RT "The sequence and analysis of Trypanosoma brucei chromosome II.";
RL Nucleic Acids Res. 0:0(2003).
KW EMBL; AB017167; AAQ15611.1; .
KW Hypochemical protein.
SQ SEQUENCE 1679 AA; 197152 MW; 3A9D4F648B99F09A CRC64;

```

Alignment Scores:

```

Pred. No.: 0.000179 Length: 1679
Score: 140.00 Matches: 216
Percent Similarity: 31.1% Conservative: 167
Best Local Similarity: 17.5% Mismatches: 409
Query Match: 1.95% Indels: 440
DB: 5 Gaps: 55

```

US-09-602-362E-22 (1-4115) x Q7YVS2 (1-1679)

```

QY 3849 CTGGCTGTGTTTC-----TCCAAAGAGTTTCTACTGCTGTGTTCTCATGAGTTT 3799
Db 777 ValIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCysAlaLysAlaCys 796
QY 3798 CTGTTTCTCTCTCTTTCTTATATGATATACGGTTTAAATGGTATTGTAATTAAT 3739
Db 797 HisTyrValPheMetTyrLysSerMetSerLeuArgVal----- 809
QY 3738 ATCTCTCTTTTCTCTTTTAGAGATGTTGCGATTCTCTCTCAAGAAATGAATATCAT 3679
Db 810 -----LeuCysAlaLysAlaCysHisTyrValPheTyrValGlnGlnHis 824
QY 3678 TGTATTC---TGCTTTGTGTGCAGCTTCTATGCGATGACTAATCTCTGGAAGCCAC 3622
Db 825 -ValIleMetCysPheMetCysLys---SerMetSerLeuArgValTyrValGlnLysHi 843
QY 3621 AATTTTGTCTAGTGAATAATTTCTGACTAGAGACTCTGCTGTTTCAGGTGTTTG-- 3564
Db 3564

```

```

Db 843 sValIleThrCys-----PheMetTyrLysSerMetSerLeuArgValLeuCy 859
QY 3563 -----TTCCATTATCTT---GTGCTTTTGATACATGTTTCGCTTCTCTCATGGA 3517
Db 859 sAlaLysAlaCysHisTyrValPheTyrValGlnGlnHisValIleThrCysPheMetCy 879
QY 3516 CACTCTTTTTCAGTGTCTCTTTGATGTT---CTGAAACCAATGATTTCTCTTAGAC 3460
Db 879 sAsn-----SerMetSerLeuCysAlaLysAlaCysHisTyrVal----- 895
QY 3459 ATCTCTCTGCTAATTCAGATAATTTTAGGTTTTCGATTTCTTTGAGCTCAGAAAGTGT 3400
Db 896 -----PheTyrValGlnGlnHisValI 903
QY 3399 GATGGAGCACTCATTGTTAATATCGTACTCTCATCAACATTCATTCTTTGCAAAACA 3340
Db 903 eMet-----CysPheMetTyrLysSerMetSerLeuArgValLeuCysGlu-- 918
QY 3339 GCATCTCTCAATGTGGAAGCAGGTTCTTACTTTTCTTGATGTCACAAATGATCATGGC 3280
Db 919 -----ThrAlaCysHisTyrValPheTyrValGlnLysHisValIleTh 933
QY 3279 TTGT-----ACAGCGAAGCCAGTCAGGATGTT 3253
Db 933 rCysPheMetTyrLysSerMetSerLeuArgValLeuCysAlaLysAlaCysHisTyrVa 953
QY 3252 GTATTCAATTTCTGCCTCTAGTTTCTTTGTCTGTTTTCCTCAATTAGAAATGAGCAT 3193
Db 953 lPheTyrValGlnGlnHisValIleMetCysPheMetCysLysSerMetSerLeuArgVa 973
QY 3192 GTTTCTCAGCTATCAGAACTTTTAGCTGC----- 3165
Db 973 lTyrValGlnLysHisValIleThrCysPheMetCysLysSerMetSerLeuArgValTy 993
QY 3164 -----CCACATATGAGATCCCTTTTAGTAATGATTCCTTTTCA-----GTTT 3121
Db 993 rValGlnLysHisValIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCy 1013
QY 3120 AGGCTCATCTAAGTTACGATCTTTCTTTAAATCTTAAGTCTCTCAAGATTTATTTTC 3061
Db 1013 sAlaLysAlaCysHisTyrValPheMetTyrLysSerMetSerLeuArgValLeu---- 1031
QY 3060 CTTTTCTCGATGTTGTTTAGTGTGCTATTTCAGTTTTCGATGCGCAATTTCTTTTCA 3001
Db 1032 -----CysAlaLysAlaCysHisTyrValPheTyrValGlnGl 1044
QY 3000 ACATGCAATTTTCATGAAGAAATAATTTTCATTTTCATGAGTTGAGAA----- 2955
Db 1044 nHisValIleMetCys-----PheMetCysAsnSerMetSerLeuCysVa 1059
QY 2954 -----ACCTATTCAAATTTCTTACACCTTCAATTCTTATCTGTATCTGAGA 2905
Db 1059 lLeuCysAlaThrAlaCysHisTyrValPheTyrValGln-----Gl 1073
QY 2904 GCCTTTCAAGTTGTTTCTACTTCTTCTATGCTGCTCTCGATTTCTTCTAATTC 2845
Db 1073 nHisValIleMetCysPheMetCysLysSerMetSerLeuArgValSerCysThrLysAl 1093
QY 2844 TTCTCAATTTTCTTTTAAATATATATCGCATTTCTCTCTCTCTCTTTGTTTAAAGTCA 2785
Db 1093 aCys-HisTyrValPheTyrValGlnLysHisValIleThrCysPheMetCysAsnSerM 1113
QY 2784 ATCTACACTCCAGACTCTTG-----TTCCCAATTAATTTTGTGTTCTCTA 2740
Db 1113 etSerLeuCysValLeuCysAlaLysAlaCysHisTyrValPheTyrValGlnGlnHisV 1133
QY 2739 CTGTGATTATTTCTTTTGTCTGAGAGTTCTTTTTCAGTA-----C 2698
Db 1133 alIleThrCysPheMetCysAsnSerMetSerLeuCysValLeuCysAlaLysAlaCysH 1153
QY 2697 AAAAACTCTTTTTCATTGTTCCATTTTCTGTACGTTTTCAGGTGTTTCAGGTGTTTGAAGTT 2638
Db 1153 is-----TyrValPheHisValGln----- 1159

```

QY 2637 CCCTGCGCTTCAACAGATGAACCTGACCAAGATTTTGTAGAGGTAGTGAATCTTCTAAT 2578
Db 1160 -----LysHisValIleThrCys-----PheMet 1168
QY 2577 TTTCACACATTTATTCCTTTCTTTTGTGTGTAGCTGGGACACACACCTCTTCTGTGA 2518
Db 1168 y-LysSer---MetSerLeuArgValLeuCysThrLysAlaCysHisTyrValPheTyrV 1187
QY 2517 AACAGTCCACGGAGACTTCAGATCCCAAGATTTTCAACCTCTCTTTGTTGATCTG 2458
Db 1187 aGln----- 1188
QY 2457 AAGGAAATCTGATCTCAATGTGTCTCAATTTTCAATTTCAATTTCAAGGTTTATTTGGAACA 2398
Db 1189 -----LysHisValIleThrCysPheMetCysAsn- 1198
QY 2397 GACTTTTTCATTTCAATGCAGGCTCGAAGCAGATGGCTCTCGGAGGCGCTCTTGAAGT 2338
Db 1199 -----SerMetSerLeuArg-----ValLeuCysAlaT 1208
QY 2337 TTGCATGCCATCAATCAAGCTTTAGTGAATAGAACTTTCAATCTCAGGGAGCCTTCA 2278
Db 1208 hrAlaCysHisTyr----- 1212
QY 2277 GAAACCTCATATCAGAGACTCTTCTATTTTCCATTTATTTATCTATTTCTTTTGTATGT 2218
Db 1213 -----ValPheTyrValGlnLysHisValIleThrCysPheMetC 1226
QY 2217 GPAGCTGGTAAACACACATCTTCTGGAAACAGCTTCACAGAGCTCCAGAAATCCCAAGA 2158
Db 1226 ys-----AsnSerMetSer----- 1230
QY 2157 ACTTCTCATAGTCTTTGTTGATTTGATGGAGTTCTCACTGCTCAATGTTGT 2098
Db 1231 -----LeuCys----- 1240
QY 2097 CATTC-----TTAATTCAGAGTTTATTTGGACAGACTTTTGAT 2059
Db 1240 isTyrValPheTyrValLysGlnHisValIleMetCysPheMetTyrLysSerMet--- 1258
QY 2058 TTCAATGGAGCTCGAAGCGAGATGGCTCTCGGAGGCGCTTGAAGTTTGCATGTCA 1999
Db 1259 -----SerLeuArgValI 1263
QY 1998 TCAATTTCTAGCTTTAGTTGGAATAGAACTTTCAATCCAGTTAGCTCAGAGACCACT 1939
Db 1263 euCysThrLys----- 1266
QY 1938 TTACAGGACCTTCTTAATTTTCCATTTATTTATTTATTTATTTTGTAGGCGCTTGG 1879
Db 1267 -----AlaCysHisTyrValPheTyrValGlnLysHisValI 1279
QY 1878 TAAACACAACTCTTCTGTGAACAGCTTCACAGAGCTCCAGTATCCCAAGATTTTCTAT 1819
Db 1279 le---ThrCysPheMetCysAsnSerMetSerLeu----- 1289
QY 1818 AGTCTTTGTTGATTTCTGATGGAGTTCTCATCTGCTCCCTCATGTTTTCATTTTCAT 1759
Db 1290 -----CysValLeuCysAlaLysAlaCysHisTyrValP 1301
QY 1758 TCCAAGGCTTATTTGGGACAGACTTTTGTATTCAGTGGAGGCTCGAAGCAGATGGCTCC 1699
Db 1301 heMet-Cys-----LysSerMetSerLeu 1308
QY 1698 CGGAGGCTCGCTTTGAAAGTTTC-----ATGCTTCATCTCTA 1660
Db 1309 ArgValLeuCysAlaLysAlaCysHisTyrValPheTyrValGlnGlnHisValIleMet 1328
QY 1659 GGCTTTAGTGAATAGAAATTTCAATCTCCGAGGTAGCTTCAGAGACCACTTTATTAGG 1600
Db 1329 CysPheMetCysLysSerMetSerLeuArgValIleThrCysHisValIleThrCys 1348

QY 1599 ACTCTCTATTTTCCATTTTATATCTATTTCTTTTGTAGTGTAGCTTTGGTAAACACAA 1540
Db 1349 PheMetCysLysSerMetSerLeuArgValIle----- 1359
QY 1539 TCCCTGTGTAAACAGTCTCCAGAGACTCTAGAAATCCCAAGAAATTTCTTATAGTCTTTG 1480
Db 1360 -----CysAlaThrAlaCysHisTyrValPheTyrValGlnLysArgValIleMetCys 1377
QY 1479 TTTGGATTCGGTGGAAACAGGATCTCTCAATGTTTGTTCATTTCTTATTCATCAAGGCT 1420
Db 1378 PheMetCysAsnSerMetSerLeuArgValLeuCysAlaThrAlaCysHisTyrValPhe 1397
QY 1419 TATTGGAAGAGTTTGTCTTCAATGGCGCTTGAAGCAGATGGCTTTTAGGAGGCTT 1360
Db 1398 TyrValGlnLysHisValIleThrCysPheMetTyrLysSerMetSerLeuArgValLeu 1417
QY 1359 TCTACTTCTTATTTATCTCATTTACTTTTATATATAGACTCAGGTATACCACTTGAATCT 1300
Db 1418 CysThrLys---AlaCysHisTyrValPheTyrVal----- 1428
QY 1299 TGCAGAACTTCAAAGAGACCCGAGAAATCAAGAAATTTCTTCACTTCTCTTTTGTGGAT 1240
Db 1429 -----GlnGlnHisValIleMetCysPheMetCysLysSer 1440
QY 1239 CTGATGGGACCTCTGATCATGCGACTTGTCTTGTAGATGATTTCTTTGTGGACATGCAAC 1180
Db 1441 ValSerLeuCysValLeuCysAlaThrAla-CysHis-TyrValPheTyrValGlnGlnH 1460
QY 1179 ATCTTAGATTTCTTTTCAAACTTTAGTTTATTTAGATGTTTACTTCTTGCACGCTCAT 1120
Db 1460 isVal-----IleMetCysPheMetCysLysSerMetCys 1471
QY 1119 CTTTACAGGTTTCTTTTCTCCATCGCTCTCTAGTCTTCTTCTTCTCCACCACTA 1060
Db 1471 erLeu-ArgValSerCysThrLysAlaCysHisTyrValPheTyrValGlnLysHisVal 1490
QY 1059 ATTCTCAGTGTCTTTTGGGACTCATATTTCCCTAGGTGTCTTCTTTTCTCCCT 1000
Db 1491 IleThr---CysPheMetCysLysIleValSer-----LeuCysValLeuCysAlaLys 1507
QY 999 GCGATCTTCTAGTCTTCTTCTGCTGCGCCGAAATTTCTCAGATGTTTCTTTCAGGAT 940
Db 1508 AlaCysHisTyrValPheMetCysLysSerVal----- 1518
QY 939 CGTAATTTCTAGTGTCTTTCTGCTGAGCTTCGAATTTCCAGATGCTTTTCTTCCAAC 880
Db 1519 -----SerLeuCys----- 1521
QY 879 ATTGAATTTCTCAGATGTTCTTCCCAAGATGCAGCTCATCAGGTGTTTTCACACAG 820
Db 1522 -----ValLeuCysAlaThrAla-----CysHisTyrValPheTyrValGln 1535
QY 819 CTTTCAGCTGTGTCAGTGTCTTTCACAGGGTGCAGCTCATCAGGTGTTTTCACCA 760
Db 1536 Lys-ArgValIleMetCysPheIleLeuLysAsnArg-ValIleMetCysPheMetTyrL 1555
QY 759 GCITTCAGGTGTCAGTGTCTTTCGCCAGG-----GTGCAGC 721
Db 1555 ysSerMetSerLeuArgValLeuCysAlaLysAlaCysHisTyrValPheTyrValGlnG 1575
QY 720 CTCATCAGGTTTCTCAGAGTTCTCTGATTTGTGATTTGTGATTTTGTATTTTATTTT 663
Db 1575 lnHisValIleMetCysPheMetCysLysSerMetSerLeuArgValSerCysThrLysA 1595
QY 662 -----TCGTATATATTC-----ATAATTTGTTTGAATGT----- 633
Db 1595 laCysHisTyrValPheTyrValGlnLysHisValIleThrCysPheMetCysLysIleV 1615
QY 632 -----GAGAAATCCA 623
Db 1615 alSerLeuCysValLeuCysAlaLysAlaCysHisTyrValPheTyrValGlnLysHisV 1635
QY 622 CAGTAACAGCATATGTTCTGCAG-----TTACTCCACTATATCTGCACAAAGAGCTCAC 569

Db	1123	HisValIleAlaLysGlyLysGlnIleAsnAsnAspLysAsnArgAsnGluAspTyrTyr	1144
QY	2639	ACTTCAA AAAAGACACTGTGAACACGTCACAGAAAATGGAACAATCAAAAAGAGTGTGTG	2698
Db	1143	ThrTyrLysAspAsnThrGluLysLysAsnGluLysHISerLeuAsnAsnTyrIle	1162
QY	2699	TACTGAAAAGAACTGTCAGNACACAAAGAATAAATCACAGTAGAGACCAAAAGTTAAA	2758
Db	1163	HisAspLysGluLysMetLysIleAsp-----AsnGluAspAsnThrAsn	1178
QY	2759	TGGGAACAGAGCTCTGCAGTGTAGATTGACTTAACCAAGAAAGACAGAGACAAATGC	2818
Db	1179	AsnLysGluAspThrProIleAsnIle-----TyrLysLysLysAsnAspAsnIle	1195
QY	2819	CGATATATTAAAGAAAATAATAGGAACAATTAGAAGAATCGAGACCAACATAGAAGAGT	2878
Db	1196	TyrIleSerLysGluCysIleThrAsnAsnAsnAsnAsnAsnAsnAsn-----	1214
QY	2879	TAGAAGTGAACACAACTTGAAGGCTCTCAGATACAGATAAGAATTGAAGCTGTAGAA	2938
Db	1215	-----AspAsnAsnGlu	1218
QY	2939	AGTAATTTGAATAGTTCCTCAACTCATGAAATGAAATTTATCTTACATCAAAATTCAT	2998
Db	1219	AsnAsnGluAsnAsnGluAsnAsnGluAsnAsnSerTyrTyrAsnAsnGlu	1238
QY	2999	GTTGAAAAGGAATTGCCATCGCAAACTGGAACTAGCCACACTAAACCAATCCAGAAA	3058
Db	1239	GlyGulysAsnThrTyrAsnGlnAsnIleSerAsnHisLeuGlnAsnLysAspLysLys	1258
QY	3059	AG-----GAAATAAATCTTTGAGGACTTAAGATTTTAAA-----GAA	3096
Db	1259	LysTyrThrAsnSerIleGluAsnGluAlaLeuIleLysAsnAsnIleLysAsnAspGlu	1278
QY	3097	AGATGCTGAACCTTAGATGACCCCTAAACT-----GAAAAGGAATCATTA	3141
Db	1279	LysTyrIleAsnTyrGlnAsnTyrLysGluArgLysIleLeuHisLysAsnAsnMet	1298
QY	3142	CTA-----AAAGGATCTCAATATGTGGCGACGTAAAGTTCTGTAGCTCAGAAAACAAT	3195
Db	1299	IleAsnLysAsnAspIleGluSerThrProGlnLeuGlnMetSerTyrIleLeuAsnAsn	1318
QY	3196	GCTCATTTCTAAATTGAGGAAAAACAGACAAAGAAACTAGAGGCAGAAAATTAATACACC	3255
Db	1319	SerGlnAsnAsnTyrSerIleAsnLysAspAsnIleIleLysSerProIleGlu-----	1336
QY	3256	ATCTCGTACTGGCTTCGCTGTACAAAGCCATGATCAATTTGTACATCAAGAAAAAGTAAGA	3315
Db	1337	-----IleValArgLysLeuLeuAsnPheAspGluLysGlnLys-----	1349
QY	3316	CTGTGTTCCACATTCAGAGATGCTGTTGTCAAAAGAAATGAATGTTGATGTGAGTAGTAC	3375
Db	1350	-----AsnIleGluGluCysIle-----GluAsn	1357
QY	3376	GATATTAAATAGTGCTCCATCACCACCTTTCTGAGCTCAAAGAAATCCAAACCTAA	3435
Db	1358	IleAsnAsnAsnAsnAsnAsnIleAsn-----AspLys-TyrProLysGluLys	1373
QY	3436	AAATTATCTCAATTAGCAGAGATGCTTAAGAGAATAATCATTGTTTCAGAACATCACAA	3495
Db	1373	AsnLysLeuLysGluAspGlnIle-----LysHis-LeuLysLysHisLeuAsnLeuAsnA	1405
QY	3496	AGAGACAACGTGAAAACAGTGTCAATGAAGGAAGCGAACACATGTATCAAAAACGACAAGA	3555
Db	1385	ysTyrAspAspMetAsnSerMetProAsnAspAsnAspAsnValHisAsnLeuAsnA	1405
QY	3556	TAAATGGAACAACCTTGAACACAGAGTCTGTAGTCAGAATATTTCACACTACAGCA	3615
Db	1405	snLeuCysMetIleSerLysGluLysAsnIleIleThrLys-----LysAsnThrThrGluLys	1424
QY	3616	AAAATTGTGGCTTCACAGCAATTAGTCATGCACATAGAAAGAGCTGCACAAACAGCAGATA	3675
Db	1424	ysTyrCysAsnIle-----PheLysAsnIleAsnAsnLysIle-----AsnCysAsnA	1440

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus n2p model

Run on: July 15, 2004, 08:45:37 ; Search time 34.7876 Seconds
(without alignments)

Title: US-09-602-362E-22

Perfect score: 7086
Sequence: 1 ctagtctatcagcaacgac.....gagtgaactccacctggaaa 4115

Scoring table: BLOSUM62

Xgapop	10.0	, Xgapext	0.5
Ygapop	10.0	, Ygapext	0.5
Fgapop	6.0	, Fgapext	7.0
Delop	6.0	, Delext	7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

```
-MODEX=frames n2p_model -DEV=xlp
-Q/cgnt2_1/USFTO_spool_p/US09602362/runat_15072004_093625_21988/app_query.fasta_1.10325
-DB=Swissprot_42 -QEXT=fastan -SUFFIX=rsp -MTNMATCH=0.1 -LOCPCL=0 -LOCPXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=us09602362@CGN.1.134 @runat_15072004_093625_21988 -NCPUS=6 -ICPU=3
-NO WMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSELOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-DEFGAEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7
```

Database : SwissProt 42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	144	2.0	3911	1	AKA9	HUMAN	Q99966 h a-kinase
2	139	2.0	1818	1	HMW2	MYCPN	P75471 mycoplasma
3	138	1.9	590	1	NUSM	TRYBB	P04540 trypanosoma
4	136	1.9	976	1	SCPI1	HUMAN	Q15431 homo sapien
5	133.5	1.9	527	1	NUEM	CAEEL	P24896 caenorhabdi
6	130.5	1.8	1433	1	REST	CHICK	P24892 caenorhabdi
7	130	1.8	547	1	NUSM	ASCUS	Q97fkl1 clostridium
8	130	1.8	1163	1	SBCC	CLOAB	P24884 ascaris suu
9	130	1.8	2230	1	GOA4	HUMAN	Q13439 homo sapien
10	130	1.8	2663	1	CENE	HUMAN	Q02224 methanococc
11	127	1.8	1005	1	RA50	METJA	Q58718 methanococc
12	125.5	1.8	2245	1	MYSL	DICDI	P54697 dictyostelli
13	125	1.8	1928	1	MYSL	YEAST	P08964 saccharomyc
14	123.5	1.7	3259	1	GOBL	YEAST	Q14789 homo sapien
15	123	1.7	313	1	NUSM	RHISA	Q09817 rhinicephal
16	121.5	1.7	1411	1	FEAL	HUMAN	Q15075 homo sapien
17	120.5	1.7	409	1	NUAM	CAEEL	P24892 caenorhabdi
18	120.5	1.7	1972	1	MYTH	RABIT	P35748 oryctolagus

RESULT 1

AKA3_HUMAN	STANDARD;	PRT; 3911 AA.
ID	AKA9 HUMAN	
AC	Q99936; Q14869; Q43355; Q94895; Q9UQH3; Q9UQ04; Q9Y6B8; Q9Y6Y2;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-MAR-2004 (Rel. 43, Last annotation update)	
DE	A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)	
DE	(PRK9) A-kinase anchor protein 450 kDa (AKAP 450) (A-kinase anchor	
DE	protein 350 kDa) (AKAP 350) (lgAKAP 350) (AKAP 120 like protein)	
DE	(Hyperion protein) (Votlak protein) (Centrosome- and Golgi-localized	
DE	PRK-associated protein) (CG-NAP).	
GN	AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.	
OS	Homo sapiens (Human)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI TaxID:9606;	

ALIGNMENTS

C	19	118.5	1.6	615	1	NUOL_BUCAP	Q8K9x7	buchnera ap
C	20	118.5	1.6	634	1	YCK3_EUGGR	P31916	euglena gra
C	21	118	1.6	486	1	NUON_BUCAP	Q8K9x5	buchnera ap
C	22	118	1.7	978	1	RA50_AQUAE	O67124	aquifex aeo
C	23	118	1.7	1727	1	ALM1_SCHPO	Q9utk5	schizosacch
C	24	117.5	1.7	1290	1	RA50_SCHPO	Q9utj8	schizosacch
C	25	118	1.6	1938	1	MYHD_HUMAN	Q9UKx3	homo sapien
C	26	115.5	1.6	469	1	NUON_BUCAL	P57264	buchnera ap
C	27	115.5	1.6	3210	1	CENF_HUMAN	P49454	homo sapien
C	28	114	1.6	1957	1	SPOF_SCHPO	P10411	schizosacch
C	29	113	1.6	494	1	NUON_BUCBP	Q89at4	buchnera ap
C	30	111.5	1.6	1790	1	USOL_YEAST	P25386	saccharomyc
C	31	111	1.6	1961	1	MYH9_RAT	Q28212	rattus norv
C	32	110.5	1.5	444	1	NUAM_JOCMI	P65424	locusta mig
C	33	110.5	1.6	845	1	SCP1_MESAU	Q60563	mesocricetu
C	34	110.5	1.6	1875	1	MLP1_YEAST	Q02455	saccharomyc
C	35	110	1.5	284	1	COX3_LEITA	P14546	leishmania
C	36	110	1.6	1427	1	REST_HUMAN	P30622	homo sapien
C	37	110	1.6	1934	1	MYH7_MESAU	P13540	mesocricetu
C	38	109.5	1.5	310	1	Y160_BUCAP	Q8K9x4	buchnera ap
C	39	109	1.5	527	1	NU2M_ACACA	Q37376	acanthamoeb
C	40	109	1.5	1505	1	SCP2_RAT	Q70608	rattus norv
C	41	109	1.5	1679	1	Y109_YEAST	P40457	saccharomyc
C	42	109	1.5	1335	1	MYH7_EIG	P79293	sus scrofa
C	43	108.5	1.5	229	1	ATP6_TRYBB	P24499	trypanosoma
C	44	108.5	1.5	736	1	NUSC_HUMAN	Q9nti4	cenothera h
C	45	108.5	1.5	1530	1	SCP2_OENHO	Q9Bx26	homo sapien

RL J. Biol. Chem. 274:17267-17274 (1999).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Kemmer W.A., Deiss S., Schwarz U.;
RT "Cloning of Hyperion";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
RC TISSUE=Gastric parietal cell;
RX MEDLINE=99115654; PubMed=9915945;
RA Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
RT Trotter K.W., Milgram S.L., Goldenring J.R.;
RA "AKAP350, a multiply spliced protein kinase A-anchoring protein
RT associated with centrosomes.";
RL J. Biol. Chem. 274:3055-3066 (1999).
RN [6]
RP SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
RC TISSUE=Lymphoblast;
RA Hinds K., Sutterer C., Becker M., Hawkins M.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
RC TISSUE=Lung;
RA Milgram S.L., Goldenring J.R., Schmidt P.H.;
RT "AKAP350: A multiply spliced family of proteins with centrosomal
RT association.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RT Tanaka A., Kotani H., Nomura N., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286 (1998).
RN [9]
RP SEQUENCE OF 17-1800 FROM N.A.
RA Wu X., Graves T., Bradshaw H.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds to type II regulatory subunits of protein kinase
CC A. Scaffolding protein that assembles several protein kinases and
CC phosphatases on centrosome and Golgi apparatus where physiological
CC events can be regulated by phosphorylation state of protein.
CC substrates. Isoform 4/Yotiao is associated with the N-methyl-D-
CC aspartate receptor and is specifically found in the neuromuscular
CC junction (NMJ) as well as in neuronal synapses explaining that its
CC role may be to organize postsynaptic specializations.
CC -!- SUBUNIT: Interacts with the regulatory region of protein kinase N
CC (PKN), protein phosphatase 2A (PP2A), protein phosphatase 1 (PPI)
CC and the immature non-phosphorylated form of PKC epsilon.
CC -!- SUBCELLULAR LOCATION: Centrosomal in many cell types and
CC cytoplasmic in parietal cells.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=6;
CC Name=1;
CC IsoId=Q99996-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
CC Name=3; Synonyms=CG-NAP;
CC IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
CC Name=4; Synonyms=Yotiao;
CC IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
CC Name=5;
CC IsoId=Q99996-5; Sequence=VSP_004108;
CC Name=6; Synonyms=AKAP350;
CC IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
CC -!- TISSUE SPECIFICITY: Widely expressed, Isoform 4/Yotiao is highly
CC expressed in skeletal muscle and in pancreas.
CC -!- DOMAIN: Kii-binding site, predicted to form an amphipathic helix,
CC could participate in protein-protein interactions with a
CC complementary surface on the R-subunit dimer.

CC -!- CAUTION: Ref.6 sequence differs from that shown due to two
CC frameshifts in positions 3782 and 3811.
CC -!- CAUTION: Ref.9 sequence differs from that shown due to four
CC frameshifts in positions 29, 1653, 1699 and 1735.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AJ131693; CAB40713.1; -;
DR EMBL; AB019691; BAA78718.1; -;
DR EMBL; AJ010770; CAA09361.1; -;
DR EMBL; AF026245; AAB86384.1; -;
DR EMBL; AF083037; AAD22767.1; -;
DR EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR EMBL; AF091711; AAD39719.1; -;
DR EMBL; AB018346; BAA34523.1; -;
DR EMBL; AC000666; AAC60380.1; ALT_FRAME.
DR Genew; HGNC:379; AKAP9.
DR MIM; 604001; -;
DR GO; GO:0005813; C:centrosome; TAS.
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr...; TAS.
DR GO; GO:0005515; F:protein binding; TAS.
DR GO; GO:0007165; P:signal transduction; TAS.
DR GO; GO:0007268; P:synaptic transmission; TAS.
DR GO; GO:0006810; P:transport; TAS.
DR GO; GO:0006810; P:transport; TAS.
KW Coiled coil; Alternative splicing; Polymorphism.
FT DOMAIN 2554 2567 PKA-RII SUBUNIT BINDING DOMAIN.
FT DOMAIN 164 914 COILED COIL (POTENTIAL).
FT DOMAIN 944 1022 COILED COIL (POTENTIAL).
FT DOMAIN 1100 1185 COILED COIL (POTENTIAL).
FT DOMAIN 1253 1280 COILED COIL (POTENTIAL).
FT DOMAIN 1336 1392 COILED COIL (POTENTIAL).
FT DOMAIN 1434 1459 COILED COIL (POTENTIAL).
FT DOMAIN 1585 1659 COILED COIL (POTENTIAL).
FT DOMAIN 1857 2455 COILED COIL (POTENTIAL).
FT DOMAIN 2544 2561 COILED COIL (POTENTIAL).
FT DOMAIN 2603 2776 COILED COIL (POTENTIAL).
FT DOMAIN 3065 3092 COILED COIL (POTENTIAL).
FT DOMAIN 3124 3470 COILED COIL (POTENTIAL).
FT DOMAIN 3587 3699 COILED COIL (POTENTIAL).
FT DOMAIN 3726 3730 POLY-LEU.
FT DOMAIN 203 292 GLN-RICH.
FT DOMAIN 321 1010 GLU-RICH.
FT DOMAIN 1846 2772 Missing (in isoform 2 and isoform 3).
FT VARSPLIC 17 28 /FTID=VSP_004102.
FT VARSPLIC 1637 1642 CLOEEL -> LATRED (in isoform 4).
FT VARSPLIC 1643 3911 Missing (in isoform 4).
FT VARSPLIC 2175 2182 /FTID=VSP_004104.
FT VARSPLIC 2175 2182 Missing (in isoform 3).
FT VARSPLIC 2175 2183 /FTID=VSP_004105.
FT VARSPLIC 2175 2183 SADTFQKVE -> Q (in isoform 6).
FT VARSPLIC 2895 2907 /FTID=VSP_004106.
FT VARSPLIC 2895 2907 VFGFVNCFTLC -> GSSPELAHSDAYOTREICSS
FT VARSPLIC 2895 2948 (in isoform 2, isoform 3 and isoform 6).
FT VARSPLIC 3901 3911 /FTID=VSP_004107.
FT VARSPLIC 3901 3911 Missing (in isoform 5).
FT VARSPLIC 3901 3911 /FTID=VSP_004108.
FT VARSPLIC 3901 3911 STTQFHAGMRR -> ALSLTTSWQHHSARPTAPLFFELILSH
FT VARSPLIC 3901 3911 SLG (in isoform 6).
FT VARSPLIC 3901 3911 /FTID=VSP_004109.
FT VARSPLIC 3901 3911 K -> Q.
FT VARSPLIC 3901 3911 /FTID=VAR_010926.
FT VARSPLIC 3901 3911 E -> Q (IN REF. 3).
FT VARSPLIC 3901 3911 M -> I (IN REF. 3).
FT VARSPLIC 3901 3911 E -> G (IN REF. 3).
FT VARSPLIC 3901 3911 M -> G (IN REF. 3).


```

Db      686  etSerGlnLysIleGluThrMetGlnPheGluLysAspAsnLeuIle-ThrLysGlnAsn 705
QY      3574 AACACGAGTCTCTAGTCAGAAATTTTCAACTACAGCAAAAATGTG-----GCT 3627
Db      706  GlnLeuIleLeuGluIleSerLysLeuLysAspLeuGlnSerLeuValAsnSerLys 725
QY      3628 TCACAGCAATTAGTCATCATCAGAAAGCTGACAAACAAAGCAGATAACATGATTC 3687
Db      726  SerGluGluMetThrLeuGlnIleAsnGlu---LeuGlnLysGluIleGluIleLeu--- 743
QY      3688 ATTCTCTGAGAGAAATGCAACACATCTCTCTAAAGAGAAAAAGAGAGAGATATTAT 3747
Db      743  ----- 743
QY      3748 AATACCATTTAAACCGTATATTCAATATGAAAAGAGAAAGAGAAAGAACTCATG 3807
Db      744  -----ArgGlnGluLysGluLys----- 750
QY      3808 AGAACACGACGTAGAAATCTTTGGAGAAACACAGACAGACTTTTACTCAGACTCATG 3867
Db      751  -----GlyThrLeuGluGlnValGlnGluLeuGln 761
QY      3868 CTAGAGCCAGCTAGCATCATCTATCTGTAATCTTACCAAGTCTGTGTACAGAT 3927
Db      762  LeuLysThrGluLeuLeuGluLysGlnMetLys-----GluLysGluAsn 776
QY      3928 ACTATTTTAGAGAAATTTCTATGTTCTCTGAGCCTACAGCATATAAATGATGCA 3987
Db      777  AspLeuGlnGlu---LysPheAlaGlnLeuGluAlaGluAsnSerIleLeuLysAspGlu 796
QY      3988 AGATTACTTGTTCGGAATTCATAGCTGCACAGGATTCCTATACCTGATGTCAGCA 4047
Db      796  YsLysThrLeuGluAspMetLeuLysIleHis---ThrProValSer---GlnGluGlu 814
QY      4048 GACTCATTCATCAACACAGAAATCGCTCTGCTACCTCCAGCCTAGTCAGAGATGGA 4107
Db      814  rgLeuIlePheLeuAspSerIleLysSerLysSerLys-----AspSerV 829
QY      4108 CCTGAAA 4115
Db      829  alTrpGlu 831

```

RESULT 2

```

ID      HMW2 MYCPN      STANDARD;      PRT; 1818 AA.
AC      P75471.
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Cytochrome high molecular weight protein 2 (Cytochrome accessory
DE      protein 2).
GN      HMW2 OR MFN310 OR MF526.
OS      Mycoplasma pneumoniae.
OC      Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
CX      NCBI_taxID=2104;
EN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 29342 / M129;
RX      MEDLINE=97105885; PubMed=9948633;
RA      Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA      Herimann R.;
RT      "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT      pneumoniae.";
RL      Nucleic Acids Res. 24:4420-4449 (1996).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 29342 / M129;
RX      MEDLINE=97252497; PubMed=9098065;
RA      Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA      Herimann R.;
RT      "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT      pneumoniae cytoskeletal protein HMW2 and cytochrome.";

```

```

RL      J. Bacteriol. 179:2668-2677 (1997).
CC      -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC      STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC      CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC      PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
CC      IN THE MYCOPLASMAL MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC      SIMILARITY).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch)
CC      -----
CC      EMBL; AB000051; AAB96174.1; -.
CC      EMBL; US9896; AAB52527.1; -.
CC      PIR; S73852; S73852.
CC      KW      Phosphite; P75471; -.
CC      FT      DOMAIN 31 880 COILED COIL (POTENTIAL).
CC      FT      Cytochrome; Structural protein; Coiled coil; Complete proteome.
CC      FT      DOMAIN 919 1607 COILED COIL (POTENTIAL).
CC      FT      DOMAIN 1644 1755 COILED COIL (POTENTIAL).
CC      FT      DOMAIN 1786 1817 COILED COIL (POTENTIAL).
CC      SQ      SEQUENCE 1818 AA; 215622 MW; 66DF4B08F0FCBC0 CRC64;
Alignment Scores:
Pred. No.: 0.000442 Length: 1818
Score: 139.00 Matches: 236
Percent Similarity: 35.48% Conservative: 200
Best Local Similarity: 19.20% Mismatches: 384
Query Match: 1.96% Indels: 409
DB: Gaps: 56
US-09-602-362E-22 (1-4115) x HMW2_MYCPN (1-1818)
QY      901 AAGTTCGACGTCAGCAGAAACACCTAGGAAATACGATCTCCAAAAAATCATCTGA 960
Db      506 LysLysGlnLysGlnAspLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 525
QY      961 GAAATTTACGGCCACCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db      526 GlnTyr-----GlnAlaAspPheGluAsnGluIleGluGluLysGlnAsnGlu 541
QY      1021 CTAGGGAAATATGAGTCCGAGAAAGAAACACCTAGGAAATACGATCTCCAAAAAAGGAGA 1080
Db      542 Leu-----PheAlaSerGlnLysSerLeuGlnLysSerPheThrGlnLysLysAsnLys 559
QY      1081 CTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1134
Db      560 GluAlaGluLeuAsn---GlnLysAlaGlnLysIleAlaGluAspTrpAlaHisLeuLys 578
QY      1135 AGTAAACATCTAATAAAGTAAAGTTTGAAGGAAATCTAAGATGTTGCGATG----- 1188
Db      579 GlnAsn-----LysHisHisAlaAspLeuGluLysPheLeu 591
QY      1189 ---CCACAAAGAGATCATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1245
Db      592 GluGlyGluPheAsnHisLeuGlnGlnLysHis-----LysLeuLeuGluAlaArg 609
QY      1246 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1305
Db      610 ThrGlnPheAspAsnArgValSerLeuLeu-----SerAlaArgPhe 623
QY      1306 AAGTGTATACCTGAGTCTATATATAAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1365
Db      623 ----- 623
QY      1366 CTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
Db      624 -----LysGlnLysGlnAlaGluLeu----- 630

```



```

Db 1236 AsnLeuSerTyLeuThrTyrLysAsnArg-AlaAspLeuSerGlnGlnLe 1255
QY 3247 GAATACACATCTGACTGGCTTCTGCTAGCAAGCCATGATCAATGTGACATCAAGAA 3306
Db 1255 u-----GlnHisLy 1258
QY 3307 AAGTAAGAACCTGCTTCCACATTGAGGAGATGCTGTTTCCAAAGAAATGAATGTGATG 3366
Db 1258 sTyAlaAsnLeuLeu-GluLeuLysGluLysLeuGlnThrAlaLysArgAlaLeuAspL 1278
QY 3367 TCAGAGTACGATTAACAATGAGTGCTCCAT-----CACACTTCTGAGCTCAA 3417
Db 1278 ySLysHisArg-----AlaIleTyrGlyLysMetAlaGlnPheValSerGluL 1294
QY 3418 AGGAAATCCAAACCTAAATATCTCAATT-----AGCAGGA 3456
Db 1294 euArgGlnGluLysGlnLeuLeuSerAlaGlnLysGlnValAspAspLysSerArgL 1314
QY 3457 GATGCTTAAGAGAAATACANTGGTTTCCAGACATC-----ACAAAGACACACCT- 3506
Db 1314 euLeuGluGlnAsnGlnArgHisLeuGlnAsnLeuSerSerGluThrLysLysArgG 1334
QY 3507 -----CAAAACAGTGTCAATGAGGAGGAGAACAC----- 3536
Db 1334 InSerLeuGluHisAspIleAsnLysPheAspGlnArgLysGluAlaValSerSerI 1354
QY 3537 --ATGTTATCAAAACGACGATAATATGGAACAAACACCTGAACAGCAGAGTCTTAGTCAG 3594
Db 1354 leuAsnSerHisLysLysLysLysGlnLysGluGlyGluLeuGlnGlyIleLeuGlnL 1374
QY 3595 AAATT--ATTCAACTACACACAAATATGCGCTC----- 3629
Db 1374 yLeuSerLeuLysLysThrGlnIleGluGlnGluPheSerLysLeuTyrGlnGlnArg 1394
QY 3630 -----ACAGCAATATGATCATGACATAGAAAGAGTGAACAAAG 3669
Db 1394 LuLysLeuAspArgGlnArgThrThrLeuSerLysLeuHisArgGluLeu-----LysA 1412
QY 3670 CAGATAACAAT-----GATATTCATTTCT 3693
Db 1412 laGlnAsnGluAlaThrAlaHisLysAsnArgGluValLeuGluLeuGluAsnTyrYL 1432
QY 3694 TGAGAGGAATGACACACATCTCCTAAAGAGAAAGAGGAGATATTAATACATACC 3753
Db 1432 yLys-GluLeuGlnArgLeuThrThrGluLysSerGluPheAspAsnLysAsnArg 1451
QY 3754 ATTAAAAACCGTATATCAATATGAAAGAGAAAGAGAAACAGAA----- 3801
Db 1452 LeuPheGluTy-PheArgLysIleArgAsnGluIleGluLysLysGluAlaHisIleLys 1471
QY 3802 -----CTCATGAGACACAGCAGTGAAGAACTCTTGGAGAAACACAGACACTTGA 3855
Db 1472 ThrValLeuGluThrGlnLysLysArgHisLeuValGlu-----ThrGluAlaValLys 1490
QY 3856 CTCACACTCATGTAGAGGCCAGTCTAGCATCACCTATG---TTGAAAATCTTACCAAG 3912
Db 1491 LeuHisLeuGlnLysGlnSerIleIleSerLysGlyGlnGluLeuLysGluIleLysGlu 1510
QY 3913 TCTGTCTCAG-----NATACTATTGA 3936
Db 1511 ArgValSerArgAspIleSerHisThrAsnLysGlnArgGluGluLeuAsnSerLeuLeu 1530
QY 3937 GAAGAAAAT 3945
Db 1531 HisGlnAsn 1533

```

RESULT 3

```

NUSM TRYBB
ID NUSM TRYBB STANDARD; PRT; 590 AA.
AC P04540;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE NADH-ubiquinone oxidoreductase chain 5 (BC 1.6.5.3).
GN ND5.
OS Trypanosoma brucei brucei.
OG Mitochondrion.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
CX NCBI_TaxID=5702;
RN [1]
RS SEQUENCE FROM N.A.
MEDLINE=85037915; PubMed=6093040;
RA Hensgens L.A.M., Brakenhoff J., de Vries B.F., Sloof P., Tromp M.C.,
van Boon J.H., Benne R.;
RT "The sequence of the gene for cytochrome c oxidase subunit I, a
frameshift containing gene for cytochrome c oxidase subunit II and
seven unassigned reading frames in Trypanosoma brucei mitochondrial
maxi-circle DNA.";
RL Nucleic Acids Res. 12:7327-7344 (1984).
RN [2]
RS SEQUENCE FROM N.A.
STRAIN=164;
RC MEDLINE=87201680; PubMed=2437452;
RA Jasmer D.P., Feagin J.E., Payne M., Stuart K.;
RT "Variation of G-rich mitochondrial transcripts among stocks of
Trypanosoma brucei.";
RL Mol. Biochem. Parasitol. 22:259-272 (1987).
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01094; CAB57807.1; -.
DR EMBL; M14820; AAB59225.1; -.
DR PIR; A04519; QOUTCS.
DR InterPro; IPR003916; NADhub_oxred5.
DR InterPro; IPR001750; Oxidored_g1.
DR Pfam; PF00361; oxidored_g1; 1.
DR PRINTS; PR01434; NADHDHGNASES.
DR Oxidoreductase; NAD; Ubiquinone, Mitochondrion.
KW OXIDOREDUCTASE; NAD; Ubiquinone; CAE231AD9DF5282 CRC64;
SQ SEQUENCE 590 AA; 71494 MW; CAE231AD9DF5282 CRC64;

Alignment Scores:
Pred No.: 0.000457 Length: 590
Score: 138.00 Matches: 107
Percent Similarity: 34.08% Conservative: 75
Best Local Similarity: 20.04% Mismatches: 162
Query Match: 1.92% Indels: 190
DB: 1 Gaps: 26

US-09-602-362E-22 (1-4115) x NUSM_TRYBB (1-590)
QY 3983 ACITTTATTTATGCTAGGCTTCAGAGAAACATGATTTCTTCTTAAATAGTATTC 3924
Db 194 ThrThrPheTyrPheLeuAsnPhePheCysMetAspTyrTyrTyrIleGluPheSerIle 213
QY 3923 TGT-----GACACAGACTTTCGTAGATTTTCAACATA 3891
Db 214 CysLeuLeuValGlyCysAlaPheThrLysSerThrGlnPheGlyLeuHisIleTrpLeu 233
QY 3890 GGTGATCTAGACTGGCTCTAGCATGAGTGAAGTCTGCTGTGTTCTTCTCCA 3831
Db 234 ProAspAla---MetGluGlyProIleProValSer-----AlaLeuIleHisAla 249
QY 3830 AAGAAGTTTCTACTG-----CTGTGTTCTCATGATTTTC-----TGTTCCTCTTC 3786
Db 250 AlaThrLeuValValCysGlyIleIleLeuLeuSerPheValTyrTrpCysPheAspPhe 269
QY 3785 TCTTTTCA-----TATTGAATATACGTTTTTAATGTTATGTAATTAATCTCCT 3732
Db 270 TrpPheSerTyrPheTyrAsnLeuIleGlyTrpSerThrLeuLeuLeuLeuMetThr 289

```

QY	3731	CTTTTCTCTTTTAGGAGATGTTGTCATTTCTCTCAAGAAATGAATATCATCATTTGTTATC	3672
Db	290	LeuCysValPhe-----TyrAsnPheAspValLysArgTyrValAlaPhe	304
QY	3671	TGCTTTTGTCGACTTTCTATGTGCATGACTAATTCTGTGAAGCCACAATTTGCT	3612
Db	305	SerThrIleCysGlnIleSerPheSerMetPheCysCysLeuCysIleAspIleTyrIle	324
QY	3611	TGTAGTTGAATAATTTCTGCTAGAGACTCTGCTGTTTCAGGTGT	3567
Db	325	GlySer-----LeuPhePheCysTyrHisMetPheTyr	335
QY	3566	-----TTGTTCCATTATCTTCTGTTTGATACATGTGTTGCTCTCTCATTTGAC	3516
Db	336	LysAlaThrLeuPheIleValLeuGlyIleTrpIleHisIlePhePheGlyLeuGlnAsp	355
QY	3515	ACTGTTTTCAGTTGTCTCTTTGATGTTCTGAACCAATGATTTCTCTTAGACATCT	3456
Db	356	-Leu-----ArgCysTyrPhePheMetTyrPheCysGlyCysValLeuAlaArgLeu	373
QY	3455	CTTGCTAATTGAGATAATTTTAGTGTTTGGATTTCTCTTTGAGCTCAGAAAGTGGTGATG	3396
Db	373	uLeuIlePheAlaIleLeu-----	379
QY	3395	GAGACATCATTTGTAATATCGTACTCTCACATCAACATTCATTTCTTTCGAACAGCAT	3336
Db	380	-----AsnSerCys-----	382
QY	3335	CTCCTCAATGTGGAAGCAGGTTCTTACTTTTCTTGATGTACAATTGATCATGGCTGT	3276
Db	382	sSer-----	383
QY	3275	ACGCGAAGCAGTCAGATGGTGATTCATTAATTTCTGCCTCTAGTTTCTTTCTCTGTTT	3216
Db	384	-----IleTrpPheLeuCys-----	388
QY	3215	TTCTCAATTTAGAATGAGCATTCGTTCTCAGCTATCAGAACITTTAGCTGCCACACATTT	3156
Db	389	-----GlyPheTyrCys-----LysAspMetLe	396
QY	3155	GAGATCCCTTTTAGTAATGATTCCTTTTCAGTTTTAGGGTCATCTAAGTTTCAGCATTTT	3096
Db	396	uLeuAlaLeuMetLeuLeuSerPheTyr-----AsnIle-IleGluPheLeuP	413
QY	3095	TCITTTAAATCTTAAGTCCTCAAGATTTATTTCCTT-----TTCC	3054
Db	413	helleSerIleIle-----PheIlePhePheThrMetIleTyrAsnTyrPheL	429
QY	3053	TGGATTGGTGTTTAGTGCGTATTCAG-----TTTTTGCAATGGCAATTCCTTTTCAACAT	2997
Db	429	euLeuPhePheLeuMetPheValPheLysCysPheCysLeuVal-----AspC	445
QY	2996	GCAATTTTCATGATATAATTTTCATTTTCATGATGAGAAACCTATTCAAAATTACTTT	2937
Db	445	ys-----LeuPheLeuLeuPheA	451
QY	2936	CTACACCTTCAANTCTTATCTTGATCTGAGAGCCTTTCAAGTTGTGTTTCACTTTCTAAC	2877
Db	451	spTyrGluCysCysLeuValTyrCysLeuIleSer-----L	463
QY	2876	TCITTTCTATGC---TGCTTCGATTCTCTAATCTTCCTAATTTTCTTTTAATATATC	2820
Db	463	eutyMetCysIleLeuSerIlePhePheIleLeuAspPheValCysIlePheValPheS	483
QY	2819	GGCATTTCCTCTCTCTCTTTCTTGTTTAAAGTCAATCTCACAT-----	2776
Db	483	erSerTyrCysValPheTrpSerPhePheLeuAsnPheTyrAsnPhePheAspIleAlaI	503
QY	2775	-----GCAGACTCTTGTTCCCATTTAACTTTTGTTTCTCTACTGTGAT	2733
Db	503	lePheValValPheLeuIleLeuSerValGlyPheLeuTyrTyrGlyCysLeuPhe-----P	522

QY	2732	TATNTCTTTTGTTCTGACAGTCTCTTTTCAGTACACAAAACCTCTTTTTCAATGGTT----	2673
Dd	522	heTyRPhPeAsnIleAaspCysileMetLeupHePTrpArgillePhePheVallellel	542
QY	2676	-----CCATTTCCTGTACGTCTTCACAGTGCCTTTTGAAGTTCCTCCT	2634
Dd	542	leLeuValValPheMeIllePheCysTrpTyRPhEValCysMeIllePheMetL	562
QY	2633	TGCCTTTTCACAAGTAGTAAGTGTACACAGATTTTGATAGCGTAGTGAATCTTTAAATTTTC	2574
Dd	562	euLeuPheValTrpAsnPheValIleTyRphe-----ArgTyRa	575
QY	2573	CACATATTTATCCCTTTCTTTTGTGTGCTGAGCTCGGTA	2536
Dd	575	snLeuLysTyRCysLeuPhePheCysileLeutrpile	587
RESULT 4			
ID	SCPI_HUMAN	STANDARD;	PRT; 976 AA.
DC	Q15431; O14963;		
DT	15-JUL-1998 (Rel. 36, Created)		
DT	15-JUN-1998 (Rel. 36, last sequence update)		
DT	15-MAR-2004 (Rel. 43, last annotation update)		
DE	Synaptonemal complex protein 1 (SCP-1 protein).		
GN	SYCP1 OR SCPI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RX	MEDLINE=97224467; PubMed=9119375;		
RA	Meuwissen R.L.J., Meerts I., Hoovers J.M.N., Leschot N.J.,		
RA	Heyting C.;		
RT	"Human synaptonemal complex protein 1 (SCPI): isolation and		
RT	characterization of the cdna and chromosomal localization of the		
RT	gene.";		
RL	Genomics 39:377-384(1997).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Testis;		
RX	MEDLINE=98037449; PubMed=9371398;		
RA	Kondoh N., Nishina Y., Tsuchida J., Koga M., Tanaka H., Uchida K.,		
RA	Inazawa J., Taketo M., Nozaki M., Nojima H., Matsumiya K., Namiki M.,		
RA	Okuyama A., Nishimune Y.;		
RT	"Assignment of synaptonemal complex protein 1 (SCPI) to human		
RT	chromosome 1p13 by fluorescence in situ hybridization and its		
RT	expression in the testis.";		
RL	Cytogenet. Cell Genet. 78:103-104(1997).		
CC	-!- FUNCTION: Major component of the transverse filaments of		
CC	synaptonemal complexes (SCS), formed between homologous		
CC	chromosomes during meiotic prophase.		
CC	-!- SUBCELLULAR LOCATION: Nuclear. In tripartite segments of		
CC	synaptonemal complexes, between lateral elements in the nucleus.		
CC	Found only towards the chromosome cores are synapsed. Its N-terminus		
CC	is found toward the centre of the synaptonemal complex while the		
CC	C-terminus extends well into the lateral domain of the		
CC	synaptonemal complex (By similarity).		
CC	-!- TISSUE SPECIFICITY: Testis.		
CC	-!- DOMAIN: Consists of an alpha-helical stretch of 700 AA residues,		
CC	flanked by N- and C-terminal globular domains. The C-terminal		
CC	domain has DNA-binding capacity (By similarity).		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		

DR	EMBL; X95654; CAAG4956.1; --		

429	LysGluValGluLeuGluLeuLysValLeuGlyGluLysGluThrLeuLeuTyr	448
2696	-----	2696
449	GluAsnLysGlnPheGluLysIleAlaGluGluLeuLysGlyThrGluGlnGluLeuIle	468
2697	-----TGTACTGAAAGAACTGTCAGAACAAAAGAAATAAAATCACA-----	2738
469	GlyLeuLeuGlnAlaArgGluLysGluValHisAspLeuGluIleGlnLeuThrAlaIle	488
2739	-----GATGAGAAC	2747
489	ThrThrSerGluGlnTyrTyrSerLysGluValLysAspLeuLysThrGluLeuGluAsn	508
2748	CAAAAGTTA---AATCGGAACAGAGCTCGAGTGTAGATTGACTTAACCAAGAAAAGA	2804
509	GlyLysLeuLysAsnThrGluLeuThrSerHisCysAsnLysLeuSerLeuGluAsnLys	528
2805	GAA-----GAGAGAAATGCGATATA-----TTAAAGAAAAAATTAGGAAGATT	2849
529	GluLeuThrGlnGluThrSerAspMetThrLeuGluLeuLysGlnGluAspIle	548
2850	AGAAGATCGAGACGACATGAAAGAGTTAGAGTGAAGCAACAACCTGGAAGCGCTCTCA	2909
549	AsnAsnAsnLysLysGlnGluGluArgMetLeuLys-----GlnIleGluAsnLeuGln	566
2910	GATACAGATAAGAAATTGAAGTCTGACAAAGTAATTTGAATAGTTTCTCAACTGAA	2969
567	GluThrGluThrGlnLeuArg-----AsnGlu	575
2970	ATGAAAATTATCTTACATGAAATTCGATGTGAAAGAGAAATTCGCATGCAAACTGGA	3029
576	LeuGluTyrValArgGluGluLeuLysGlnLysArgAspGluValLysCysLysLeuAsp	595
3030	ATAGCCACACTAAACCCAAAT-----CCAGAAAAGAAAATAAATC	3071
596	LysSerGluGluAsnCysAsnAsnLeuArgLysGlnValGluAsnLysAsnLysTyrIle	615
3072	TTTGAGACCTTAAGATTTTAAAGAAAAAGTCTGAACTTAGATGACCTAAACTGAAA	3131
616	---GluGluLeuGlnGluAsnLysAlaLeuLysLys-----Lys	628
3132	GGAAATCACTAAAGGGATCTCATATGTGGCAGCTAAAGTT-----	3176
629	GlyThrAlaGluSerLysGlnLeuAsnValTyrGluIleLysValAsnLysLeuGluLeu	648
3177	---CTGATAGCTGAGAAACAATGCTCATTTCTAAAT-----GAGGAAAA	3218
649	GluLeuGluSerAlaLysGlnLysPheGlyGluIleThrAspThrTyrGlnLysGluIle	668
3219	CAGACAAAAGAACTAGAGCAGAAAATTGAATACACATCCGCTGCGTTCGCTGTACA	3278
669	GluAspLysLysIleSerGluGluAsnLeuLeuGluValGluLysAlaLysValIle	688
3279	AGCCATGATCAATGTGCATCAAGAAAAAGTA---AGAACCTGCTTCCAC---ATTGAG	3332
689	AlaAspGluAlaValLysLeuGlnLysGluIleAspLysArgCysGlnHisLysIleAla	708
3333	GAGATGCTGTTTGCAGAAAAATCAATGTTGATGTGAGATACGATATTAAACAATCATG	3392
709	GluMetValAlaLeuMetGluLys-----	716
3393	CTCATCACCAAC-----TTTCTG	3410
717	---HisLysHisGlnTyrAspLysIleIleGluGluArgAspSerGluLeuGlyLeuTyr	735
3411	AGCTCAAG-----GAAATCCAAAAC	3431
736	LysSerLysGluGlnGluGlnSerSerLeuArgAlaSerLeuGluIleGluLeuSerAsn	755
3432	CTAAAA-----ATTATCTCAATTAGCAGAGATGCTTAAGAGAAATACATTCGTTTCAG	3485
756	LeuLysAlaGluLeuSerValLysGlnLeuGluIleGluArgGluGluLysGlu	775

Qy	3486	AA	CATCAACAAGAGACACACCTGAAACACAGTGTCAATGAAGGAAGCGACACATGTTATCAA	3545
Db	776	Lys	LeuLysArgGluAlaLysGluAsnThrAlaThrLeuLysGluLysLysAspLysLys	795
Qy	3546	AAC	GACACAGATAATGGAACAAACACACCTGAA---CAGCAGAGTCTCTAGTCAGAAAATTATT	3602
Db	796	Thr	GlnThrPheLeuLeuGluThrProGluLeuLeuTriPlyLysLeuAspSerLysAlaVal	815
Qy	3603	TC	AACTACAAGCAAAATTTGGCTTCACAGCAATTAGTCATGCACATAGAAAG---	3656
Db	816	Pro	SerGlnThrValSerArgAsnPheThrSerValAspHisGlyIleSerLysAspLys	835
Qy	3657	---	CTGACACACAAAGCAGATAACAATGA---	3685
Db	836	Arg	AspTyrLeuThrThrSerAlaLysAsnThrLeuSerThrProLeuProLysAlaTyr	855
Qy	3686	TC	ATTTCTTGAGAGGAAATGCAACATCTCTTAAAGAGAAAAAGAGAGATATTAAATT	3745
Db	856	Thr	Val-----LysThrProThrLysProLysLeuGlnGlnArgGluAsnLeu---	871
Qy	3746	ACA	TACCATTTAAAAACCGTATATTCAATATCAAAAAGAGAAAGAGA-----	3793
Db	872	Asn	IleProLeuGluGlu-----SerLysLysLysArgLysMetAlaPheGluPhe	888
Qy	3794	---	-----AACGAAACTCATGAGAACAA	3814
Db	889	Asp	IleAsnSerAspSerSerGluThrThrAspLeuLeuSerMetValSerGluGluGlu	908
Qy	3815	GC	AGTAGAAACTTCTTTGGAGAAACACAGACAGACTTTACTCAC-----AC	3862
Db	909	Thr	LeuLysThr-LeuTyrArgAsnAsnAsnProProAlaSerHisLeuCysValLysTh	928
Qy	3863	TC	ATGCTAGAGCGCAGCTTAGCATCAC	3890
Db	928	rPro	LysLysAlaProSerSerLeuThr	937
RESULT 5				
NUSM_CAEEL				
ID	NUSM_CAEEL	STANDARD;	PRT;	527 AA.
AC	P24896;			
DT	01-MAR-1992	(Rel. 21, Created)		
DT	01-NOV-1995	(Rel. 32, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	NADH-ubiquinone oxidoreductase chain 5	(EC 1.6.5.3).		
GN	N05 OR WICE.35.			
OG	Caenorhabditis elegans.			
OC	Mitochondrion.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;			
OC	Rhabditidae; Peloderiinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2;			
RX	MEDLINE=92201635; PubMed=1551572;			
RA	Okimoto R., Macfarlane J.L., Clary D.O., Wolstenholme D.R.;			
RT	"The mitochondrial genomes of two nematodes, Caenorhabditis elegans			
RT	and Ascaris suum,"			
RL	Genetics 130:471-498 (1992).			
RN	[2]			
RP	SEQUENCE OF 1-120 FROM N.A., AND VARIANT LEU-94.			
RC	STRAIN=AB1, AB2, Bristol N2, CB4852, CB4853, CB4854, CB4855, CB4856,			
RX	CB4857, CB4858, KX314, PB303, PB306, RW7000, and TR403;			
RC	MEDLINE=22531585; PubMed=12644560;			
RA	Denver D.R., Morris K., Thomas W.K.;			
RT	"Phylogenetics in Caenorhabditis elegans: an analysis of divergence			
RT	and outcrossing,"			
RL	Mol. Biol. Evol. 20:393-400 (2003).			
CC	-1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			

QY 820 CTGTGGGAAAACACCTGATGAGGCTGCATCTTGGTGAGGAACATCTGACAAATTCAT 879
Db MetValGluLysHisLysThrGluLeuGlu
QY 880 GTTGAGAAAGGACATCTGGAAGTTCGAAGCTCAGCAGCAAAAACACCTAGGAATACG 939
Db SerLeuLysHisGlnGlnAspAlaLeuThrGluLysLeuGln 632
QY 940 ATCTGCGAAAACATCTGAAATTTACGGGCCAGCAAGAGACCTAGAAAGATCGC 999
Db ValLeuLysGlnGln--TyrGlnThrGluMetGluLysLeuArgGluLys 648
QY 1000 AGGGAAGAAAAGACACACCTAGGAAA-----TATGATCCCGGAAA 1044
Db CysGluGlnGluLysGluThrLeuLysAspLysGluLeuPheGlnAlaHisLe 668
QY 1045 GAACACTGGAATACCTGGCGGAGAAAAGAGACCTAGGAAGCCTAGGAGAAA 1104
Db GluGluMetAsnGluLysThrLeuGluLysLeuAspValLysGlnThr-- 684
QY 1105 AGAAACCTGTAAAGATGATGCGTGCAGAGTAACATCTAATAACTAAAGTTTGA 1164
Db GluLeuGluSerLeuSerGluLeu-----SerGluValLeuLys 698
QY 1165 AAGGAA-----ATTAAGATGTTGCATGTCCACAAAAG 1197
Db AlaArgHisLysLeuGluGluLeuSerValLeuLysAspGlnThrAspLysMetLys 718
QY 1198 AATCATCTACAAAGCAAGTGCATCAGAGTCCCATCGATCGAACAAGGAGAT 1257
Db GlnGluLeuGluAlaLysMetAspGluGlnLysHisHisGlnGlnValAsp 737
QY 1258 GAAGAATATCTTGATCTCGGTCTCTTTG-----AAGTTCTGCAAGATTC 1305
Db Ser--IleIleLysGluHisGluValSerIleGlnArgThrGluLysAlaLeuLysAspG 757
QY 1306 AAGTGGTATACCTGAGTCTATATAAAAGTAATAGATAATAAAG-- 1355
Db IlnIleAsnGlnLeuGluLeuLeuLysGluArgAspLysHisLeuLysGluHisGlnA 777
QY 1356 -----TAGAAGCCT 1365
Db laHisValGluAsnLeuGluAlaAspIleLysArgSerGluGlyGluLeuGlnGlnAla 797
QY 1366 CCTAAAAGCCATCTGCTTCAAG----- 1388
Db erAlaLysLeuAspValPheGlnSerTyrGlnSerAlaThrHisGluGlnThrLysAla 817
QY 1389 -----CCGCCATTGAAGCAAACTCTTCCAAATAAGCCT-- 1424
Db yrGluGluGlnLeuAlaGlnLeuGlnLysLeuLeuAspLeuGluThrGluArgIleL 837
QY 1425 --TTGAATGAAGAATGAACAAACATTCGAAGCA-----GATCCGTGT-- 1463
Db euLeuThrLysGlnValaGluValGluAlaGlnLysLysAspValCysThrGluLeuA 857
QY 1464 --TCCCACCGAATCCAAACAAAGACTATAGAAATCTT----- 1502
Db spAlaHisLysIleGlnValGlnAspLeuMetGlnGlnLeuGluLysGlnAsnSerGluM 877
QY 1503 -----GGGATTTCTAGAGTCTCTGAGACTGTTTACAGAAGGATTTGTTTACCAAGCTA 1557
Db etGluGlnLysValLysSerLeuThrGlnValTyrGluSer--LysLeuGluAspGlyA 896
QY 1558 CACATCAAAAGATAGATAAAAAATGGAATAATAGAA--GAGTCCCTAATAAGTGGT 1614
Db snLysGluGlnGlnThrLysGlnIleLeuValGluLysGluAsnMetIleLeuGlnM 916
QY 1615 CTTCTGAGGCTACTCGAATGAATTTCTATTCCACTAAAGCCTAGAAATTCAGACAT 1674
Db etArgGluGlyGlnLysLys-----GluIleGluLeuThrG 929

QY 1675 GCAAACTTTCAAGCGAGCCTCGGGAAGCCATCTGCTTCGAGCCTCCACTGAAATCAA 1734
Db lnlLysLeuSerAlaLys-----GluAspSerIleHis----- 939
QY 1735 AGTCTCTCCAAATAAGCCTTGGATGAAAAATGAAAAACATCGAGCGCATGAGAACTC 1794
Db -----lleLeuAsnGluGluTyrGluThrLysPheLysAsnGlnGluLysMetGluL 958
QY 1795 CCATCAGATCCAAACAAAGGACTATAAGAAATTCGGGATACTGGAGTCTCTCTGTAGAC 1854
Db ysValLysGlnLysAlaLysGluMetGlnGluThrLeu----- 970
QY 1855 TGTTTCACAGAGGATGTGTTTACCAGGCTGCGCTCAAAAAGATAGATAAAATAATG 1914
Db -----lysLysLysLeuAspGlnGluAlaLysLeuLysLys----- 983
QY 1915 GAAATTAAGAGGTCCTGTTAAAGTGTCTTCTGAGCTAACTGGGAATGAAAGTTCT 1974
Db --GluLeuGluAsnThrAlaLeuGluLeuSerGlnLysGlu-----LysGlnP 999
QY 1975 ATTCCAACTAAAGCCTAGATTCATGACATGCATCAAACTTCAAGACGCGCTCCGAGAGC 2034
Db heAsnAlaLysMetLeuGluMetAla-GlnAlaAsn--SerAlaGlyIleSerAspAla 1017
QY 2035 ATCTGCTTCGAGCCTCCATTGAAATCAAAAGTCTGTCCAAATAAACCTTGAATTAAGA 2094
Db ValSer-----ArgLeuGlu 1022
QY 2095 ATCAACAAACATTCAGGAGATGAGAACTCCCATCAATCCAAACAAAGGACTATGAGAA 2154
Db ThrAsnGlnLys-GluGlnIleGluSerLeuThrGluValHisArgArgGluLeuAsnAs 1042
QY 2155 AGTCTCTGGATTCGAGTCTCTGTAGACTGTTCCAGAGGATGCTGTTTACCAGGC 2214
Db pValIleSerIleTrpGlu-----LysLysLeuAsnGlnGlnAl 1055
QY 2215 TACATCAATAAAGATAGATATAAATAATGGAATAATAGAAAGTCTCTGATTAATGAGTT 2274
Db aGluGluLeuGlnGlu-----lleHisGluIleGlnLeuGlnGluLys 1069
QY 2275 TTCTGAA-----GGCTCCCTCAGAAATGAAAGTCTTATTTCCACTAAAGCCTTGAATTTGAT 2328
Db sGluGlnGluValAlaGluLeuLysGlnLysIleLeu-----LeuPh 1083
QY 2329 GGCATGCAAACTTTCAAAGCAGCGCTCCCGAGAGCCATCTGCTTCGAGCCTCATTTGAA 2388
Db eGlyCys----- 1085
QY 2389 TGAAGTCTGTCCAAATAA-----CCTTGAATTAAGAA-----TG 2427
Db -GluLysGluGluMetAsnLysGluIleThrTrpLeuLysGluGluGlyValLysGlnAs 1105
QY 2428 AACAACTTGACACAGATCAGATTTCCCTTCAGAAATCAAAACAAAGAGTTGAGAAAT 2487
Db pThrThrLeuAsnGluLeuGlnGlnLysGlnLysSerAlaHisValAsnSerle 1125
QY 2488 TCTGGATTCGAGTCTCCGTGAGTCTTTTCCAGAGAGGATGGTGTGTACCCAGCTTAC 2547
Db uAla-GlnAspGluThrLysLeuLysAlaHisLeuGlu-----LysLeuG 1140
QY 2548 ACACAAAAAGAAAGATATAATAGTGGAAATTAAGAGATTCACTAGCCTATCAAAATCT 2607
Db luValAspLeuAsnLysSerLeuLysGluAsn----- 1150
QY 2608 TGGTACAGTTTCATCTTGTGAAAGCGCAAGGAACTTCAAAAAGACACTGTGACACAGTACA 2667
Db -----ThrPheLeuGlnGlnGluGlnLeuValGluLeuLysM 1162
QY 2668 GGAATGCAACATGAAAAAGAGTTTGTGTACTGAAAAAGAACTCTCAGACAAAAGA 2727
Db eLeuAlaGluAspLysArgLysVal-----SerGluLeuThrSerLysLeuL 1179
QY 2728 AATAAATCACAGTAGAAGCAACCAAAAGTTA-----AATGGCAACAGAGTC 2772

822	Db	eLysThrLeuHisMetAspPheGluGlnLysTyrLys-----MetValle	837
945	Qy	GCAAAAAAC-----ATCTGAGAAATTTACGGCCACGACAAAGAA-----	984
837	Db	uGluGluAsnGluArgMetAsnGlnGluIleValAsnLeuSerLysGluAlaGlnLysPh	857
985	Qy	-----GACCTAGAGA	995
857	Db	eAspSerSerLeuGlyAlaLeuLysThrGluLeuSerTyrLysThrGlnGluLeuGlnG	877
996	Qy	TCGCGAGGAGAAAGAACACACCTAGCGAAATATGACTCCCGAAAGAAACACTGAG	1055
877	Db	uLysThrArgGluValGlnGlnGluArgLeuAsnGluMetGluGlnLeuLysGluGlnLeuG	897
1056	Qy	AAATTACGTGGGAGAAAAGAACACTAGAGACGCATCGGAGAGAAAAGAAACCTGT	1115
897	Db	uAsn-----ArgAspSerProLeuGlnThrValGluArgGluLysThrLeuI	913
1116	Qy	AAAGATGGATCGTGGCAAGAGTAACATCTAATAAACTAAAGTTTGGAAAGAAATCTA	1175
913	Db	e-----ThrGluLysLeuGlnGlnThrLeuGluGluValLysThrLeuThrGlnGluLysAs	932
1176	Qy	AGATGTTGGCATGTCACAA-----AAGAACTCATCTACAAAGCAAGTG	1217
932	Db	pAspLeuLysGlnLeuGlnGlnSerLeuGlnIleGluArgAspGlnLeuLysSerAspI	952
1218	Qy	CCATGATCAGAGTCCCATCAGATCCAAACAAACGGGAAGT-----	1257
952	Db	eHisAspThrValAsnMetAsnIleAspThrGlnGlnGlnLeuArgAsnAlaLeuGluSe	972
1258	Qy	-----GAAGAATAATCTTTGATTTCTCGGGTCTCTTGAAGTTCTGCAAGATTCAAGTGGTA	1313
972	Db	rLeuLysGlnHisGlnGluThrIleAsnThrLeuLysSerLys-----	986
1314	Qy	TACGTGAGTCTATATATAAAAGTAATGAGATAAATAAGAGTAGAAGCTCCTCTAAAG	1373
987	Db	-----IleSerGluGluValSerArgAsnLe	995
1374	Qy	CCATCTGCCTTCAAGCGCCATTGAAAGCAAAACTCTTTCCAA-----	1416
995	Db	uHisMetGluGluAsnThrGlyGluThrLysAspGluPheGlnGlnLysMetValGlyI	1015
1417	Qy	-----ATAAGCCTTGAATGAAGATGAACAAACATTG-----	1449
1015	Db	eAspLysLysGlnAspLeuGluAlaLysAsnThrGlnThrLeuThrAlaAspValLysAs	1035
1449	Qy	-----	1449
1035	Db	pAsnGluIleIleGluGlnGlnArgLysIlePheSerLeuIleGlnGluLysAsnGluLe	1055
1450	Qy	-----AAGCAGATCCGTGTTCC-----ACCGA	1472
1055	Db	uGlnGlnMetLeuGluSerValIleAlaGluLysGluGlnLeuLysThrAspLeuLysG	1075
1473	Qy	ATCCAAACAAAGACTATTAAGAAAATCTTGGATTCTAGAGTCTCTGGAGACTGTTTA-	1531
1075	Db	uAsnIleGluMetThrIle-GluAsnGlnGlnLeuArgLeuGluIleAspGluLeuL	1095
1532	Qy	-----CAGAAGGATTCGTTTACCA-----GGCTACA	1559
1095	Db	ysLysGlnGlnGluIleValAlaGlnGluLysAsnHisAlaIleLysLysGluGlyGluL	1115
1560	Qy	CATCAAAAAGATAGATAAAAAATGGAATAAGAGAGTCCCTATAAAGTGGTCT---	1616
1115	Db	eUsrArgThrCysAspArgLeuAlaGlu-ValGluGluLysLeuLysGlnLysSerGln	1134
1617	Qy	-----TCTGAAGGCTAC	1628
1135	Db	GlnLeuGlnGluLysGlnGlnGlnLeuLeuAsnValGlnGluLwMetSerGluMetGln	1154
1629	Qy	CTCGGAATGAAATTTCTATTCACCTAAAGCCCTAGATTGAAGACATGCAAACTTTCA---	1685
1155	Db	LysLysIleAsnGluIleGluAsnLeuLysAsnGluLeuLysAsnLysGlnLeuThrLeu	1174

Qy	996	TCCAGGAGGAGAAAAGAACACACACCTAGGGAATAATGACTCCCGAAAAAGAAACACTGAG	1055
Db	877	uLysThrArgGluValGlnGluArgLeuAenGluMetGluGlnLeuLysGluGlnLeuG1	897
Qy	1056	AAATTCAGTGGGAGAAAAGAAAGACATGACAGACGCATGGCAGAAAAAGAAACCCCTGT	1115
Db	897	uAen-----ArgAspSerProLeuGlnThrValGluArgGluLysThrLeu11	913
Qy	1116	AAAGATCGATGCGTGCAGAGTAGTAACTAACTAACTAAAGTTTGTGAAAAGGAAATCTA	1175
Db	913	e-----ThrGluLysLeuGlnGlnThrLeuGluGluValLysThrLeuThrGlnGluLysAs	932
Qy	1176	AGATGTTGCATGTCACAA-----AAGAATCATCTACAAAGCAAGTG	1217
Db	932	pAspLeuLysGlnLeuGlnLysSerLeuGlnLleGluArgAspGlnLeuLysSerAsp11	952
Qy	1218	CCATGATCAGAGTCCCATCAGATCCAAACAAGGAGAT-----	1257
Db	952	eHisAspThrValAsnMetAsnLleAspThrGlnGluGlnLeuArgAsnAlaLeuGluSe	972
Qy	1258	----GAAGAATATTCCTTGATTCGCGGTCTCTTGAAGTCTGCAAGATTCGAAGTTCAGTGGTA	1313
Db	972	rLeuLysGlnHisGlnGluThrLleAsnThrLeuLysSerLys-----	986
Qy	1314	TACCTGAGTCTATATATAAAAGTAAATGATAAATAAGAAAGTAGAAGAGCTCCTCTAAAAG	1373
Db	987	-----LleSerGluGluValSerArgAsnLe	995
Qy	1374	CCATCTGCCTTCAAGCGCCCATTAAGCAAAACTCTTTCCAA-----	1416
Db	995	uHisMetGluGluAsnThrGlyGluThrLysAspGluPheGlnGlnLysMetValGly11	1015
Qy	1417	-----ATAAGCCTTTGAATGAAGATGAACAAACATTC-----	1449
Db	1015	eAspLysLysGlnAspLeuGluAlaLysAsnThrGlnThrLeuThrAlaAspValLysAs	1035
Qy	1449	-----	1449
Db	1035	pAsnGluLleGluGlnGlnArgLysLlePheSerLeuLleGlnGluLysAsnGluLe	1055
Qy	1450	-----AAGCAGATCCGTGTTCC-----ACCGA	1472
Db	1055	uGlnGlnMetLeuGluSerValLleAlaGluLysGluGlnLeuLysThrAspLeuLysG1	1075
Qy	1473	ATCCAAACAAGACTATTAAGAAATCTTGGGATTCCTAGAGTCTCGAGACTGTGTTA-	1531
Db	1075	uAsnLleGluMetThrLle-GluAsnGlnGluThrLeuArgLeuGluGlyAspGluLeu1	1095
Qy	1532	-----CAGAAGGATTCGTGTACCAA-----GGCTACA	1559
Db	1095	ysLysGlnGlnGluLleValAlaGlnGluLysAsnHisAlaLleLysLysGluGlyGlu1	1115
Qy	1560	CATCAAAAAGATAGATAAAAAATGGAAAAATAGAAAGTCCCTTAATAAAGTGGTCT---	1616
Db	1115	euSerArgThrCysAspArgLeuAlaGlu-ValGluGluLysLeuLysGlnLysSerGln	1134
Qy	1617	-----TCTGAAGGCTAC	1628
Db	1135	GlnLeuGlnGluLysGlnGlnGlnLeuLeuAenValGlnGluLwMetSerGluMetGln	1154
Qy	1629	CTCGGAATGAATTTCTATTCCACTAAAGCCCTAGATTTGAAGACATGCAAACTTTCA---	1685
Db	1155	LysLysLleAsnGluLleGluAsnLeuLysAsnGluLeuLysAsnLysGlnLeuThrLeu	1174

[illegible]

1686 QY -----AAGCGAGCTCCGGGAGCCATCTCTCTCGAGCTCCACTGAAATCAAAAGTCT 1739
 1175 Db GluHisMetGluThrGluArgGluLeuGluAlaGlnHisLeuAsnGluAsnTyrGluGlu 1194
 1740 QY GTCCCAAT-----AAGCTTTGGAATGAAATGAAACATGGAGGCGAG 1784
 1195 Db VallysSerIleThrLysGluArgLysValLeuLysGluLeuGlnLysSerPheGluThr 1214
 1785 QY ATGAGAATCCCATCAGA-----ATCCCAACAAAG 1814
 1215 Db GluArgAspHisLeuArgGlyTyrIleArgGluIleGluAlaThrGlyLeuIleThrLys 1234
 1815 QY GACTATAAGAAATCTCGGATCTGAGTCTCTGTAGACTGTTTCACAGAAGGATTGTG 1874
 1235 Db GluGluLeuLysIle----- 1239
 1875 QY TTTACCAAGGCTGCGCTCAAAAGNATAGATAAATAATGAAATTAGAAGGTCCTG 1934
 1240 Db ---AlaHisIleHisLeuLysGluHisGlnGluThrIleAspGluLeuArgArgSerVal 1258
 1935 QY TTAAGTGTCTCTCGAGCTAACTGGGAATGAAAGTTCTATTCCAACTAAAGCCTAGAA 1994
 1259 Db SerGluLysThrAlaGln-----IleIleAsnThrGlnAspLeuGlu 1272
 1995 QY TTGATGACATGCAACTTCAAGCAGGCTCCGAGAGCCATCTGCTTCGAGCCTCAT 2054
 1273 Db LysSerHisThrLysLeuGlnGluIleProValLeuHisGluGlnGluLeuLeu 1292
 2055 QY TGAATCAAAAGTCTGCAAAATCAACCTTGAATTAAGATGAA----- 2099
 1293 Db ProAsnValLysLysValSerGluThrGlnGluThrMetAsnGluLeuGluLeuThr 1312
 2099 QY ----- 2099
 1313 Db GluGlnSerThrLysAspSerThrThrLeuAlaArgIleGluMetGluArgLeuArg 1332
 2100 QY -----CAAACATCAGGAGCATGAGAACTCCATCAATCCAA-ACAAGGACTATGAG 2152
 1333 Db LeuAsnGluLysPheGlnGlnSerGlnGluIleLysSerLeuThrLysGluArgAsp 1352
 2153 QY AAAGTCTTGGATCTCGAGTCTCTGAGACTGTTTCCAGAAGGATGGTCTTTACCCAG 2212
 1353 Db AsnLeuLysThrIleLysGluAlaLeu-----Glu 1362
 2213 QY GCTACACATCAA-----AAGANATAGATAAATAATGCG-----AANATGAGAGTCT 2260
 1363 Db VallysHisAspGlnLeuLysGluHisIleArgGluThrLeuAlaLysIleGlnGluSer 1382
 2261 QY CTGATAATGAGCTTTTCTGAAGCTCCCTGAGATGAAAGTTCTATTCCACTAAAGCCTT 2320
 1383 Db -----GlnSerLysGlnGlnGlnSerLeuAsnMetLysGluLysAspAsnGluThrThr 1400
 2321 QY GAATTCATGCGCATGCAAACTTTCAAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCTG 2380
 1401 Db LysIleValSer-GluMetGluGlnPheLysProLysAspSerAlaLeuLeuArgIleGlu 1420
 2381 QY CATTGAATGAAAGTCTGTTCCAAATAAACCCTTCGAATTAAGATGAACAACATTGAGA 2440
 1420 Db uileGluMet-----LeuGlyLeuSerLysArgLeuGlnGluSe 1433
 2441 QY CAGATCAGATTTCCCTTCAGAAATCAAAACAAAGA-----AGTTGAGAAAT 2488
 1433 Db rHisAspGluMetLysSerValAlaLysGluLysAspAspLeuGlnArgLeuGlnGluVal 1453
 2489 QY CTGGGATTCGAGTCTCGTGGATCTGTTTCACAGAAGGATGGTGTGTACCCAGCTACA 2548
 1453 Db lLeuGlnSerGlu----- 1457
 2549 QY CACAAAAGAAAGGATAAATAGTGGAAATPAGAATTCAGCTCATCAAAATCTT 2608
 1458 Db -----SerAspGlnLeuLysGluAsn-----IleLysGluIle 1468

2609 QY GGTACAGTTTCATCTTGTGAAAGCGAAGGAACTT----- 2642
 1468 Db eValAlaLysHisLeuGluThrGluGluGluLeuLysValAlaHisCysCysLeuLysGlu 1488
 2643 QY -CAAAAGACACTGTGAACACGTACAGCAAAATGGAACAATCAAAAGAGTTTGTGTGAT 2701
 1488 Db uGlnGluGluThrIleAsnGluLeu-----ArgValAsnLeuSe 1501
 2702 QY TGAAGAAGAACTGTCTAGAACAAAGAAATAATCACTAGTAGAGAACCAAGTTAAATGG 2761
 1501 Db rGluLysGluThrGluIleSer-----ThrIleGlnLysGlnLeuGluAlaIle 1517
 2762 QY GAACAAGAGTCTGAGTGTAGATTGACITTAACCAAGAAAGAGAGAGAGAGA----- 2813
 1517 Db eAsnAspLysLeuGlnAsnLysIleGlnGluLysTyrGluLysGluGlnLeuAsnIle 1537
 2814 QY -----AATGCCGATATATTAAAGAAATAATAGGAAGAAAT 2848
 1537 Db elysGlnIleSerGluValGlnGluAsnValAsnGluLeuLysGlnPheLysGluHisArg 1557
 2849 QY TAGAAGATCGAGACGACATAGAAAGAGTTAGAACTGAACACAACTTGA----- 2900
 1557 Db gLysAlaLysAspSerAlaLeuGlnSerIleGluSerLys---MetLeuGluLeuThrAs 1576
 2901 QY -AGGCTCTCAGATACAAAGATAAAGATTGAAGGTGTAGAAAGTAATTTGAATAGTGTCTC 2959
 1576 Db nArgLeuGlnGluSerGlnGluGluIleGlnIle-----MetIleLysGlu 1591
 2960 QY AACTCATGAAATGAAATTTTCTTACATGAAATTCATGTTGCAATGTGAAAGAGTAATGCCA-- 3017
 1591 Db uLysGluGluMetLysArgValGlnGluAlaLeuGlnIleGluArgAspGlnLeuLysGlu 1611
 3018 QY ---TCRAAACTGGAATAGCCACACATCCACCAATCCAGGAAAGGAAATAAATCTT 3073
 1611 Db uAsnThrLysGluIleValAlaLysMetLysGluSer-GlnGluLysGluThrGlnPheL 1631
 3074 QY TGAGG-----ACTTAAGATTTTAAAGAAAGATG----- 3102
 1631 Db euLysMetThrAlaValAsnGluThrGlnGluLysMetCysGluIleGluHisLeuLysG 1651
 3103 QY -----CTGAATTTAGTACCCCTGAACTGAAAGGAAATCAATCAT 3142
 1651 Db luGlnPheGluThrGlnLysLeuAsnLeuGluAsnIleGluThrGluAsnIleArgLeu 1671
 3143 QY TAAAGGGATC----- 3153
 1671 Db hrGlnIleLeuHisGluAsnLeuGluMetArgSerValThrLysGluArgAspAspL 1691
 3154 QY -----TCAATATGTGGCAGCTAAAGTTCTGATGCTGAGAAACAATGCTCATTAAT 3208
 1691 Db euArgSerValGluGluThrLeuLysVal-----GluArgAspGlnLeuLysGluAsnL 1709
 3209 QY TGAGGAAAAACACAGAAAGAACTAGAGGCAAGAAATTCATATACACATCTCTGCTGCT 3268
 1709 Db euArgGluThrIleThrArgAspLeuGluLysGlnGluGluLeuLysIle----- 1725
 3269 QY TCCTGTACAGCCATGATCAATTTGACATCAAGAAAGTAAGAACCTGCTTCCACAT 3328
 1726 Db -----ValHisMetHisLeuLysGluHisGlnGluThrIleAspLysL 1740
 3329 QY TGAGGAGATGCTGTTTGCAGAAAGAAATGAATGTTGATGTGAGAGTACGATATTAACAATG 3388
 1740 Db euArgGlyIleValSerGluLys----- 1750
 3389 QY AGTCTCCAT-----CACCATTTCTGAGCTCAAGAGGAAATCCAAAC 3432
 1750 Db lIleSerAsnMetGlnLysAspLeuGluHisSerAsnAspAlaLeuLysAlaGlnAspL 1770
 3433 QY TAAATAT-----ATCTCAATTAGCAGGAGATGTCTTAAGAGAA-----A 3471
 1770 Db euLysIleGlnGluLeuArgIleAlaHisMetHisLeuLysGluGlnGlnGluThrI 1790
 3472 QY TACATTGGTTTCAGAACATCAACAAGAGACAACTGTAAGAAACAGTGTCAATGAGGAGCG 3531

```

Db      1790 leAspLysLeuArgGlylleValSerGluLysThrAspLysLeuSerAsnMetGlnLysA 1810
QY      3532 AACACATGTATCAAAACGACAGTAATGGAACAAACACCTGAACAGAGTCTCTAGT 3591
Db      1810 spLeuGluAsnSerAsnAlaLysLeuGlnGluLysIleGlnGlu-LeuLysAlaAsnGlu 1829
QY      3592 CAGAAATATTTCACATACAGCAAAATTCGCTTCACAGCAATAGTAGTCACATA 3651
Db      1830 HisGlnLeuIleThrLeuLysLysAspValAsnGlnThrGlnLysValSerGluMet 1849
QY      3652 GAAAGCTGACACAAACAGCAGATA----- 3675
Db      1850 GluGln---LeuLysLysGlnIleLysAspGlnSerLeuThrLeuSerLysLeuGluIle 1868
QY      3676 -----ACATGATATTCTTCGAGGAAATGCAACATCTCC 3717
Db      1869 GluAsnLeuAenLeuAglGlnGluHisGluAsnLeuGluMetLysSerValMet 1888
QY      3718 TAAAGAGAAAAAGAGGAGATATTAATTAACATACCATTTAAAAACCGTATATCAATAT 3777
Db      1889 LysGluArgAsp-----AsnLeuArgVal---GluGluThrLeuLysLeu 1903
QY      3778 GAAAGAGAGAAAGAGAAACAGAACTCATGAGACAGCAGTAGAACTCTTTGGAGAA 3837
Db      1904 GluArgAspGlnLeuLysGluSerLeuGlnGlnThrLysAlaArgAspLeuGluIleGln 1923
QY      3838 ACAACAGACAGACTTTACTCACACTCATGCTAGAGCCAGTCTAGCATCACCTATGTTG 3897
Db      1924 GlnGluLeuLysThrAlaArgMetLeuSerLysGluHisLysGluThrValAspLysLeu 1943
QY      3898 -----AAATCTTACCAAGTCTGTGTCAGAACTACTTTTGAAGAAATTCATGT 3951
Db      1944 ArgGluLysIleSerGluLysThrIleGlnIleSerAspIleGln-LysAspLeuAspLys 1963
QY      3952 TTCTTCTCGAAGCTCAGCATCAATAA-----AGTGTGAGATTAAGTCTT 3999
Db      1963 sSerLysAspGluLeuGlnLysLysIleGlnGluLysLysLysLeuGlnLeu 1983
QY      4000 CCGAATGCAATAGCTGCACAGGATTCCTCCATTCCTGATGTCAGAGACTCAT---TC 4056
Db      1983 uArgValLys-----GluAspValAsnMetSerHisLysLys 1995
QY      4057 AATCAACACAG-----AATCCGCTCTGC 4078
Db      1995 sIleAsnGluMetGlnLeuLysLysGlnPheGluProAsnTyrLeuLys 2012

RESULT 11
RA50_METJA
ID_RA50_METJA STANDARD; PRT; 1005 AA.
AC_Q58718;
DT_16-OCT-2001 (Rel. 40, Created)
DT_16-OCT-2001 (Rel. 40, Last sequence update)
DT_28-FEB-2003 (Rel. 41, Last annotation update)
DE_DNA double-strand break repair rad50 AtPase.
GN_RAD50 OR M1322.
OS_Methanococcus jannaschii.
OC_Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC_Methanocaldococcales; Methanocaldococcus.
OX_NCBI_TaxID=2190;
RN_[1]
RP_SEQUENCE FROM N.A.
RC_STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX_MEDLINE=96337999; PubMed=8688087;
RA_Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA_Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA_Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA_Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA_Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA_Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA_Cotton M.D., Roberts K.M., Hirst M.A., Kaine B.P., Borodovsky M.,
RA_Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus

```

```

RT_jannaschii";
RL_Science 273:1058-1073 (1996).
CC_-1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC_rad50/mre11 complex possesses single-strand endonuclease activity
CC_and ATP-dependent double-strand-specific exonuclease activity.
CC_Rad50 provides an ATP-dependent control of mre11 by unwinding
CC_and/or repositioning DNA ends into the mre11 active site (By
CC_similarity).
CC_-1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC_-1- SIMILARITY: Belongs to the SMC family. RAD50 subfamily.
CC-----
CC_This SWISS-PROT entry is copyright. It is produced through a collaboration
CC_between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC_the European Bioinformatics Institute. There are no restrictions on its
CC_use by non-profit institutions as long as its content is in no way
CC_modified and this statement is not removed. Usage by and for commercial
CC_entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC_or send an email to license@isb-sib.ch).
CC-----
DR_EMBL; U67572; AAB99331.1; -.
DR_PIR; A64465; A64465.
DR_TIGR; M1322; -.
DR_HAWAP; MF_00449; -.
DR_InterPro; IPR003593; AAA_ATPase.
DR_InterPro; IPR003439; ABC transporter.
DR_InterPro; IPR007517; Rad50_zn_hook.
DR_InterPro; IPR003405; SMC_C.
DR_InterPro; IPR003395; SMC_N.
DR_Pfam; PF04423; Rad50_zn_hook; 1.
DR_Pfam; PF02483; SMC_C; 1.
DR_Pfam; PF02463; SMC_N; 1.
DR_Smart; SM00382; AAA; 1.
DR_KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT_NP_BIND 32 39 ATP (By SIMILARITY).
FT_DOMAIN 138 849 COILED COIL (POTENTIAL).
SQ_SEQUENCE 1005 AA; 119387 MW; 98BBB48173E789F3 CRC64;

Alignment Scores:
Pred. No.: 0.00478 Length: 1005
Score: 127.00 Matches: 192
Percent Similarity: 33.94% Conservative: 146
Best Local Similarity: 19.28% Mismatches: 291
Query Match: 1.79% Indels: 368
DB: 1 Gaps: 41

US-09-602-362E-22 (1-4115) x RA50_METJA (1-1005)
QY 1402 CAAACTCTTCCAAATAGCCTTTGAATGAGAATGAACAAACATTGAAGCAGATCCGT 1461
Db 123 ArgAsnMetPheLeuAsnSerIleTyrIleLysGlnGlyLeuAlaLysPheLeuSer 142
QY 1462 GTTCCACCAGATCCAAACAAAGGACTATAGAAAATTTCTGGGATTCTAGAGTCTCTGG 1521
Db 143 LeuLysProSerGluLysLeuGluThrValAlaLysLeuLeuGlyLysAspGluPheGlu 162
QY 1522 AGACTGTTTACAGAGGATGTGTTTACAGAGCTACATCAAAAGAAATAGATAAAA 1581
Db 163 Lys-CysTyrGlnLys-----MetGlyLeuLysGluTyrGluLysAr 178
QY 1582 AATGAAAATAGACAGAGTCCCTAATAAAGTGTCTTCTGAAGGCTACCTCGGAATGAAT 1641
Db 178 gluGlu----- 180
QY 1642 TTCTATTCCATTAAGCCTTAAGATTAAGACATGCAAACTTTCAAGCGAGCTCCGGA 1701
Db 181 -----ArgIleGluGlyLeuAsnTyrLysGluAsn----- 191
QY 1702 AGCCATCTGCTTCGAGCCTCCACATGAATCAAACTGTCTCCAAATAGAGCTTGAATG 1761
Db 192 -----TyrGluLysGluLeuLysAsnLysMetSer-----GlnLeuGlu 205
QY 1762 AAAAAATGAAAAC---ATGGAGGCAGATGAG-----AACTCCCATCAGAAATCCAAACA 1812

```


QY	1168	GAATCTTAAGATGTTGCATGTCACAAAAAGAAATCATCTCAAAAGCAAGTGCCTCATGATCAG	1221
Db	1048	euGlnLeuSerGluCysLeuSerLysLeuGluGlnThrGlnLeuAspHisSer-	1067
QY	1228	AGGTCCCATCAGATCCAAACAAGGGAAGATCAAGAAATATTCTTTGATCTCGGCTCTCTT	1287
Db	1068	-----SerLysLeuAsnLysLysLeuGluLysAspLeuSerAspGlnHisAspSerI	1085
QY	1288	TGAAGTTCTCCAAAGATTCAAGTGGTATACCTGAGTCTTATATAAAAAGTAATAGAGATAA	1347
Db	1085	leGlu-----LysLeuGln-	1093
QY	1348	ATPAGAGTAGAAGCCCTCTCAAAGCCATCTGCCCTTCAAGCCGCCATTGAAGCAAAAC	1407
Db	1093	sn--GluThrGlnGlnLeuGlnGlnPheLysGlnGlnSerGluLeuSerLysSerLy	1112
QY	1408	TCCTTCCAAATAAGCCCTTGAATGAAGAAATGAACAAACATTGAAGCAGATCCGTGTTCCTC	1467
Db	1112	sLeuSerLysThrThrGlnGlnLeuAspPheAsnLysGln-GluPheAspArgLeuSerG	1132
QY	1468	ACCGAATCCAAACA-	1481
Db	1132	lnGluArgAspThrAspAsnThrAsnGlnLeuGluIleGlnGlnLeuLysLysAlaA	1152
QY	1482	-----AAGCATTATAAGAAAAATCTTGGGATTTCTAGAGTCTCTCGAGACTG	1527
Db	1152	snSerThrLeuGluGluAspTyrPheSerLeuSerGlyIleArgAspAsnLeuGlu-	1170
QY	1528	TTTACAGAAGATGTGTTTACCAGGCTACACATCAAAAAGAAATAGATAAAAAATGGA	1587
Db	1171	-----ArgGlnValLeu-	1177
QY	1588	AAATAGAGAGTCCCTAATAAGTGGTCTCTGAGGCTACCTCGGAATGAAATTTCTAT	1647
Db	1177	rgAspGluAsnGlnLeuIleLysGluArgLeuasp-	1188
QY	1648	TCCACTAAAGCTTAGAATTGAAGCATCCAAACTTTCAAAGCAGCCTCCGGGAAGCCAT	1707
Db	1189	-----SerLeuGlyGlnGlnSer-SerGlnPheGlnSer-----GlyAlaAla	1202
QY	1708	CTGCTTCGAGCCTCCACTGAA-----ATCAAAAGTCTGTCCCAATAAGCCTTGGAATG	1761
Db	1203	LeuGluLysGlnGlnLeuGluGlnLeuValGlnGlnSerGluGlnLeuIleLysLeu	1222
QY	1762	AAAAATGAAAAACATGAGGCGAGATGAGAACTCCCATCAGAATCCAAACAAAGCACTATA	1821
Db	1223	SerSerGluLysLeuGlySerGluGluGluAlaLysLysGlnIleAsnGlnLeuGluLeu	1242
QY	1822	AGAAAATCTCGGATACTGGAGTCTCTGTAGACTGTTTTCAGNAGAGATTGCTCTTTACCA	1881
Db	1243	Glu-----LeuThr	1245
QY	1882	AGGCTCGCTCAAAA-----AGAATAGATAAAATAATGCAAAATTAGAAAGGTCCTCTGTAA	1938
Db	1246	AspHisLysSerLysLeuGlnIleGln-	1254
QY	1939	AGTGGTCTCTGAGGCTAACT-----GGGAATGAAAGTCTTATTCACATAAGCCCTAG	1992
Db	1255	-----LeuGlnLeuThrGlnGlnSerAsnGluLys-IleLysLysLeuLysGlyLy	1271
QY	1993	AATTGATGACATGCAAACTTCAAAGCAGGCGCTCCCGAGAGCCATCTGCCTCGAGCGCTCC	2052
Db	1271	sLeuGluGluTyrGlnAspGluLysLysGlnLeuGlnGln-----GluLeuGluArgGII	1289
QY	2053	ATTGAAATCAAAGTCTGTCCAAATAAACCTTGGAAATTAGAATGAACAACATTGAGGC	2112
Db	1289	eLysGlnSerLys-----GlnSerValGluAs	1298
QY	2113	AGATGAGACTCC-----CATCAATCCAAACAAGAGCTATCAGAGAAAGTTCTTGGGAT	2166
Db	1298	pGluArgAsnSerLeuIleThrGlnLeuThrThrValLysPheGluSerThr-----	1315
QY	2167	TCTGGAGTCTCTGAGACTGTTTCCAGAAGGATGGTGTTCACCGAGGTACACATCAAAA	2226

Db	1316	 -----GlnValSerThrAsnValSerHisGlnLeu	1322
Qy	2227	AGAATAGATAAATAATGGAAA-----ATPAGAAGAGTCTCTGATGAATAGGTTTCTGA	2280
Db	1325	sGluYsileThrThrLeuLysSerThrileGluGlu-----LeuAsnLysSerIle	1342
Qy	2281	AGCTCCCTCGAGAAATGAAAGTTCTATCCACTAAAGCCITGAATTGATGCCATGCACAACT	2340
Db	1342	eGluYsLeuGln-----	1346
Qy	2341	TTCAAAAGCAGCCCTCCCGAGAGCCATCTGCTCGAGCCTGCATTGAAATGAAAAGTCTGT	2400
Db	1347	-AlaGluGlnLysAsnLysAspGlu-----	1355
Qy	2401	TCCAATAAACCTTGGAATTAAAGATGAACAACATTGACAGACATCAGATTTCCCTTCAG	2460
Db	1356	-----IleArgLysileGlnPheGluLeuAs	1364
Qy	2461	AATCAAAACAAAGAGTTGAGAAAATTCTGGCATTTCTGAAGTCTCCGTGCACTGTTTCA	2520
Db	1364	nAspGlnLysGln-----	1368
Qy	2521	CAGAAGGATGGTGTGATCCAGAGCTACACAAAAAGAAAGATAAATAGTGGAAAATT	2580
Db	1369	-----GlnPheThrArgGlnThrLys-----	1377
Qy	2581	AGAAGATTCACTAGCCCTATCAAAATCTTGGT-----ACAGTTCAAT-	2620
Db	1377	eSerAspLeuGlnSerGlnGlnSerIleAspArgProLysSerGluIleThrIleHisSe	1397
Qy	2621	-CTTGTCAAAAGCAAGGAACTTCAAAAACACTGTGAACACGTACAGAAAAT---GG	2676
Db	1397	rLeuGluuArgThrAsnGluThrLeuLysSerAspPheGluArgValGlnGlnSerLeuLys	1417
Qy	2677	AACAATGAAAAGAGTTTGTGTACTGAATAAGAACTGTCAGA-----ACAAAAGAAAT	2730
Db	1417	sGlnGlnGluArgAspCysGlnGlnTyrlYsAspThrIleAsnArgLeuGlnGlnGluVa	1437
Qy	2731	AAATCAGTAGAGAACCAAAAGTTAAATCGGAACAGAGCTCTGCAGCTAGATTCGACTT	2790
Db	1437	lYsGlnLeuThrGlnLeuLysGluArgPheGluAsnGluPhePheValalalYsGluG	1457
Qy	2791	AAACCAAGAAAAGAGAGAGAGAAATCCGATATATTTAAAGAAAATTTAGGAAGAATTA	2850
Db	1457	nAsnSerAsnGlnThrGlnGlnSerVal---TyrlYsGluValThrThrGlnMetGl	1476
Qy	2851	GAAGATCGAGACAGCATAGAAGAGTTAGAGTGAACACACA-----CTTGAAAGGC	2904
Db	1476	nGlnAsn-GlnSerArgIleGluArgGluLeuGluLysGlnHisIleThrArgI	1496
Qy	2905	TCTCAGATACAAATAGAAATTGAAGGTGTAGAAAGTAATTTGAATAGGTTTCTCAACT	2964
Db	1496	leAspAspGluArgAspGluLeuLys-----LysGlnLeuThrGlnGlnGlnGlnH	1514
Qy	2965	ATGAAATG-----AAAATTATCTTACATGAAT-----TGCATGTTGAAA	3006
Db	1514	isGluGlnSerSerThrGlnLeuLeuLeuAlaGlnAsnGluLeuGluArgLeuArgLys	1534
Qy	3007	AGGAATTGCCATGCAAA-----	3024
Db	1534	ysGluLeuLysTyrlYsGluArgGlyHisGluThrSerLysGlnGlnAspGlnPheAsnM	1554
Qy	3025	CTGGAATPAGCCACACTAAACACCAATCCAGAGAAAAGGAAAATAATCTTTGAGCACTTAA	3084
Db	1554	etGluIleGlnSerLeuArgIleThr-AsnAsnAspGlnLeuLysSerLeuGln-----	1571
Qy	3085	GATTTTAAAGAAAGAGTCTGAACTTAGATGAC-----	3120
Db	1572	AspTyrlGluGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys	1591
Qy	3121	AAAACTGAAAAGGAATCATTTACTAAAAAGGGATCTCAATATGTGGGGCAGCTTAAAGTTCTGA	3180

[illegible]

```
Db 1048 ysLeuValThrLeuGluLysAsnCys-----AsnIleAlaM 1060
QY 3496 ACAGACACAGTGAACACAGTCTCAATGAGGACCGACACATGTATCAACAGACAGA 3555
Db 1060 etSerArgLeuGlnSerLeuValThrGluAsnSerAspLeuArgSerLysAsnGluAsnP 1080
QY 3556 TAAATGGAACAA-----ACACTGACACAGCAGTCTCTAGTCAAGAAATTATTT 3603
Db 1080 heLysLysGluLysAlaAlaLeuAsnAsnGlnLeuLysAsnLysGluSerGluLeuLeuL 1100
QY 3604 CAATACAGCAAAATTTGGTCTTCACAGCAATGATGTCATGACATAGAAAGTGCAAA 3663
Db 1100 ysMetLysGluLys-----IleAspAsnHisLysLysGluLeuAlaThrP 1115
QY 3664 CAAAGCAGATAACAATGATATTCATTCTTTGAGAGGAATGCAACATCTCTCTAAAG 3723
Db 1115 heSerLysGlnArgAspAspAlaValSerGluHisGlyLysIleThrAlaGluLeuLysG 1135
QY 3724 AGAAATAA-----AGGAGATATTAATTTACAAATACCAT---TTAAAAACCGTATATT 3771
Db 1135 luthrArgIleGlnLeuThrGluLysLysSerAsnTyrGlnLysIleLysGluLutyrS 1155
QY 3772 CAATATGAAAGAGAGAGAGAAACAGAACTCATGAGACACAGCAGTAACTCTTTT 3831
Db 1155 er-----AsnPhcGlnArgGluThrLysGluGlnGlnLysLysLysLysLysLysLys 1172
QY 3832 GGAGAAACACAGACAGACTTTACTCACATCTCATGCTAGAGCCAGCTAGACATCA--- 3888
Db 1173 ValGluSerLeuAsnAspSerLysLysLysLysLysLysLysLysLysLysLysLysLys 1191
QY 3889 ---CCTATGTTGAAATCTTACCAAGAGTGTGTGTCACAGAAATCTATT 3933
Db 1192 lleserLeuAsnGlnTyrLeuAsnLysArgIleSerGlyAsnSerVal 1207

RESULT 14
GOBL_HUMAN
ID GOBL_HUMAN STANDARD; PRT; 3259 AA.
AC Q14789; Q14398;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Golgi autoantigen, golgin subfamily B member 1 (Giantin) (Macroglolin)
DE (Golgi complex-associated protein, 372-kDa) (GCP372).
CN GOLGB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94187728; PubMed=7511208;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
RT protein (giantin).";
RL Mol. Cell. Biol. 14:2564-2576 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94257116; PubMed=8198703;
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,
RA Renz M.;
RT "Macroglolin -- a new 376 kD Golgi complex outer membrane protein as
RT target of antibodies in patients with rheumatic diseases and HIV
RT infections.";
RL J. Autoimmun. 7:67-91 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=95100974; PubMed=7802676;
RA Souda M., Misumi Y., Fujiwara T., Nishioka M., Ikehara Y.;
RT "Molecular cloning and sequence analysis of a human 372-kDa protein
RT localized in the Golgi complex.";
RL Biochem. Biophys. Res. Commun. 205:1399-1408 (1994).
```

```
CC -!- FUNCTION: May participate in forming intercisternal cross-bridges
CC of the Golgi complex.
CC -!- SUBUNIT: Homodimer; disulfide-linked.
CC -!- SUBCELLULAR LOCATION: Golgi; rheumatoid arthritis and in the
CC autoimmune disease Sjogren's syndrome.
CC -!- SIMILARITY: Belongs to the golgin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75304; CAAS3052.1; -
DR EMBL; D25542; BAA05025.1; -
DR FIR; A56539; A56539.
DR FIR; I52300; I52300.
DR Genew; HGNC:4429; GOLGB1.
DR MIM; 602500; -
DR GO; GO:000139; C:Golgi membrane; TAS.
DR GO; GO:0005795; C:Golgi stack; TAS.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0007030; P:Golgi organization and biogenesis; TAS.
KW Golgi stack; Antigen; Coiled coil; Transmembrane.
FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 3236 3256 POTENTIAL.
FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
FT DOMAIN 48 593 COILED COIL (POTENTIAL).
FT DOMAIN 677 1028 COILED COIL (POTENTIAL).
FT DOMAIN 1062 1245 COILED COIL (POTENTIAL).
FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
FT DOMAIN 1828 3185 COILED COIL (POTENTIAL).
FT DOMAIN 2420 2423 POLY-GLU.
FT DOMAIN 2993 2996 POLY-SER.
FT CONFLICT 1 39 MISSING (IN REF. 3).
FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).
FT CONFLICT 1765 1765 D -> G (IN REF. 3).
FT CONFLICT 2950 2950 H -> D (IN REF. 3).
SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Alignment Scores:
Pred. No.: 0.0118 Length: 3259
Score: 123.50 Matches: 213
Percent Similarity: 34.91% Conservative: 155
Best Local Similarity: 20.21% Mismatches: 326
Query Match: 1.74% Indels: 363
DB: 1 Gaps: 45

US-09-602-362E-22 (1-4115) x GOBL_HUMAN (1-3259)
QY 1383 TTCAAGCCGCCATTGAAGCAAACTCTTCCA-----AAT 1418
Db 1158 TrpLysProGluLeuGluGluLysIleLeuAlaLeuGluLysGluLysGluLeuGln 1177
QY 1419 AAGCCTTTGAATGAGAATGACAAACATTGAGCAGATCCGTTCCACCAGATCCAA 1478
Db 1178 LysLysLeuGlnGluAlaLeuThr-SerArgLysAlaIleLeuLysLysAlaGlnGlu 1197
QY 1479 ACAAGAGACTATAAGAAATCTCTTG-----GGATTCTAGAGTCTCTGGAG 1523
Db 1197 sgluArgHisLeuArgGluGluLeuLysGlnGlnLysAspTyrAsnArgLeuGln 1217
QY 1524 ACTGTTTACAGAA-----GGATTGTTTACCAAGCTACA 1559
Db 1217 uGlnPheAspGluGlnSerLysGluAsnGluAsnIleGlyAspGlnLeuArgLeuGln 1237
QY 1560 CATCAAAAAGATAGATAAAAAATGGAATAATGAGAGAGTCCCTAATAAAGTGGTCTCT 1619
Db 1237 nllglnValArgGluSerIleAspGlyLys-----Le 1248
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 09:09:52 ; Search time 60.1449 Seconds
(without alignments)
13162.486 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086
Sequence: 1 ctgtctatcacagcaacgac.....gactggactccactggaaa 4115

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=cgn2.1/USPNO.spool_p/US09602362/runat_15072004_093626_22015/app_query.fasta_1.10325
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCUT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -WAITRX=Blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USR=US09602362 @CGN 1 1 248 @runat_15072004_093626_22015 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
C 1	148.5	2.1	1363	2 T15653
C 2	146	2.0	590	1 S34960
C 3	144	2.0	1642	2 T08880
C 4	143	2.0	445	2 E22845
C 5	139	2.0	1818	1 S73852
C 6	138.5	1.9	502	1 I30010
C 7	138	1.9	590	1 Q0UTCS
C 8	136.5	1.9	443	2 B26696
C 9	136	1.9	1613	2 D90129
C 10	134	1.9	421	2 T44511
C 11	133.5	1.9	527	2 S26037
C 12	133	1.9	426	2 T44522
C 13	132	1.9	1979	2 C71622
C 14	131.5	1.8	288	2 S36954

C 15	130.5	1.8	288	2 S36953	cytochrome-c oxida
C 16	130	1.8	528	2 S26025	NADH2 dehydrogenas
C 17	130	1.8	1163	1 G97236	ATPase involved in
C 18	130	1.8	2663	1 S28261	centromere protein
C 19	129.5	1.8	1231	2 S70553	chromosome-associa
C 20	128.5	1.8	620	2 A59932	cytochrome C-type
C 21	128.5	1.8	918	2 T40030	hypothetical prote
C 22	128.5	1.8	1410	1 A57013	early endosome ant
C 23	127	1.8	1005	2 A64465	hypothetical prote
C 24	127	1.8	2954	2 T14156	kinasin-related pr
C 25	126.5	1.8	288	2 S36955	cytochrome-c oxida
C 26	126	1.8	1187	2 T20544	hypothetical prote
C 27	125.5	1.8	2245	2 T18278	myosin heavy chain
C 28	125	1.8	1928	2 S46773	myosin heavy chain
C 29	123.5	1.7	1312	2 T30845	probable DNA repai
C 30	123.5	1.7	3259	1 A58539	giantin - human
C 31	123	1.7	313	2 T11153	NADH2 dehydrogenas
C 32	123	1.7	3225	2 I52300	giantin - human
C 33	122.5	1.7	531	2 T11074	NADH2 dehydrogenas
C 34	122.5	1.7	3187	2 JCS837	364K Golgi complex
C 35	121.5	1.7	509	2 F71526	hypothetical prote
C 36	121	1.7	980	2 E71606	hypothetical prote
C 37	120.5	1.7	409	2 S26033	NADH2 dehydrogenas
C 38	120.5	1.7	1972	1 A41604	myosin heavy chain
C 39	118	1.6	513	2 E71683	NADH2 dehydrogenas
C 40	118	1.7	978	2 A70387	conserved hypothet
C 41	118	1.7	1173	2 T43527	sp8 protein - fise
C 42	118	1.7	1727	2 T50073	myosin-like coiled
C 43	117.5	1.6	478	2 T11318	NADH2 dehydrogenas
C 44	117.5	1.6	547	2 T33437	hypothetical prote
C 45	116.5	1.6	437	2 S34959	NADH2 dehydrogenas

ALIGNMENTS

RESULT 1

T15653

Hypothetical protein C27D6.2 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C/Accession: T15653

R/Ding, H

submitted to the EMBL Data Library, July 1995

A/Description: The sequence of C. elegans cosmid C27D6.

A/Reference number: Z18383

A/Accession: T15653

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1363 <DIN>

A/Cross-references: EMBL:U23179; NID:G726424; PID:G726426; PIDN:AAC46722.1; CESP:C27D6.2

A/Experimental source: strain Bristol N2

C/Genetics:

A/Gene: CESP:C27D6.2

A/Introns: 76/3; 149/3; 215/2; 287/3; 353/2; 434/3; 507/3; 613/3; 690/3; 785/1; 806/3; 8

Alignment Scores:

Pred. No.:	1.39e-05	Length:	1363
Score:	148.50	Matches:	200
Percent Similarity:	33.53%	Conservative:	136
Best Local Similarity:	19.96%	Mismatches:	325
Query Match:	2.07%	Indels:	342
DB:	2	Gaps:	51

US-09-602-362E-22 (1-4115) x T15653 (1-1363)

QY	3046	GGTGTAGT---GTGGCTATTCAGTTTTCAT-----GGCAATTCTCTTTTCAAC	2999
Db	28	IlePheSerThrLeuAlaValProAlaLeuPheIlePheLeuLeuLysGlnIlePhePro	47
QY	2998	ATGCATTTTCATGTGAAGATATATTTTCATTCATGAGTTGAGAACCTATTCAATTACT	2939
Db	48	LeuProPheHisGlyAsnIleLysPheMet-----	57


```
Db 334 e-----TyrLysAlaThrLeuPheIleValLeuGly--Val 345
QY 3111 TAAGTTCAGCATCTTTCTTTTAA-----3088
Db 346 TrpIle-----HisLeuPhePheGlyLeuGlnAspValArgCysTyrPhePheIleTyrPhe 364
QY 3087 -----ATCTTAAGTCCTCAAGAGATTATTTTCTCTTTTCTCGGATTGGTGTATTAGT 3037
Db 365 CysGlyCysValLeuAlaAlaMetLeuLeuValPheAlaLeuLeuAsnSerCysSerLeu 384
QY 3036 TGGCTATTCCAGTTTGGCATGGCAATCTCTTTTCAACATGCATGCATTCATGTAAGATAA 2977
Db 385 TrpPheLeu-----CysGlyPheTyr-----391
QY 2976 TTTTCATTTCATGAGTTGAGAAACCTATTCAAAATTACTTTCTACACCTTCAATTCTTATC 2917
Db 392 -----CysLysAspLeu-----LeuLeuCysThr-----399
QY 2916 TTGATATCGAGAGCCTTCAAGTTGTGTCTCACTTCACTTCACTTCTTATGCTGCTCTCGA 2857
Db 400 LeuMetLeuValSerPheHisPheIleLeuGluPhe---LeuPheValCysIle-----416
QY 2856 TTCCTCTAATCTCTCAATTTTCTTTTCTTAATATATGCGCATTTCTCTCTCTCTTTCT 2797
Db 417 -----PhePheIlePhePheThrValIleTyrAsnTyrPheLeuLeuPhePheLeu 433
QY 2796 TGG-----TTTAAGTCAATCTACATCGACACTCTTGTTCCTCCATTTTAACCTTTTGTCT 2743
Db 434 CysPheValPheLysCysPheCysLeuAlaAsp---CysLeuPheLeuLeuPheAspPhe 452
QY 2742 CTACTGTGATTATTTCTTTCTGTCGACACTTCTTTTTCAGTACAAACTCTTTTTC 2683
Db 453 GluCysCysLeuIleTyrCysPheLeuGlyLeuTyrMetCysPheIleLeuIlePhePhe 472
QY 2682 ATTGTTCATTTCCTGTACGTGTTCACAGTGTCT 2648
Db 473 ValIleAsp--PheLeuTyrIlePheValPheSer 483

RESULT 3
T08880
NM_001001: receptor-binding protein yotiao - human
C:Species: Hmo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: T08880
R:Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A:Title: Yotiao, a novel protein of neuromuscular junction and brain that interacts with
A:Reference number: Z16511; PMID:9482789
A:Accession: T08880
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1642 <LIN>
A:Cross-references: EMBL:AF026245; NID:g2623067; PIDN:AB86384.1; PID:g2623068
C:Genetics:
A:Map position: 7q21-22
C:Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction; skeletal muscle

Alignment Scores:
Pred. No.: 3,97e-05 Length: 1642
Score: 144.00 Matches: 175
Percent Similarity: 39.36% Conservative: 143
Best Local Similarity: 21.66% Mismatches: 252
Query Match: 2.03% Indels: 240
DB: 2 Gaps: 39

US-09-602-362E-22 (1-4115) x T08880 (1-1642)

QY 1989 CTAGATTGATGATGACATGCAAACTTCAAGAGCGGCTCCGAGAGCCATCTCGCTTCGAGC 2048
Db 169 LeuGluMetMet-----GluSerGluLeuAlaGlyLys 179
QY 2049 CTCCTTGAATCAAAAGTCTGTCGCAATTAACCTTGAATTAAGAAATGAACAAACATG 2108
```

```
Db 180 GlnHis-GluIleGluGluLeuAsnArgGluLeuGluGluMetArgVal-----195
QY 2109 AGGCAGATGAGAACTCCCATCAAAATCCAAACAAAGAGCATATGAGAA---AGTTCTTGGGA 2165
Db 196 -----ThrTyrGlyThrGluGlyLeuGlnGlnGluPheG1 209
QY 2166 TTCTGGAGTCTCTGATAGACTGTTTCCAGAGGATGGTGTATTTTACCAGGCTACACATCAA 2225
Db 209 uAlaAlaIleLys-----GlnArgSpGlyIleIleThrGlnLeuThrAla-- 224
QY 2226 AAGAATAGATAAAATAATGMAAATAGAGAGTCTCTGATAAT-----2268
Db 225 -----AsnLeuGlnGlnAlaArgGluLysAspGluThrMetArgGluPheLe 241
QY 2269 -GAGTTTTCGAGGCTCCCTCGAAGATGAAGTTCTATTTCACATAAGCCCTTGAATTGA 2327
Db 241 uGluLeu-ThrGluGlnSerGlnLysLeuGlnIleGlnPheGln-----GlnLeuG 258
QY 2328 TGGCATGCAAACTTTCAGAGCAGGCTCCGAGAGCCATCTGCTTCGAGCCTGCAATGAA 2387
Db 258 lAlaSerGluThrLeuArgAsnSerThrHisSerThrAlaAlaAspLeu-LeuGln 277
QY 2388 ATCAAAAGTCTCTTCCAAATTAACCTTGGAAATTAAGATGAACACATTTGAGACAGATCA 2447
Db 278 AlaLysGln-----GlnIle-----LeuThrHisGlnGlnGlnLeuGluGln 292
QY 2448 GATTTCCTTCAGATCAAAACAAAGAGGTTGAGAAATTTCTGGATTCTGAAGTCTCC 2507
Db 293 AspHisLeuLeuGluAspTyrGlnLysLysLysGluAspPhe-----ThrMetGln 309
QY 2508 GTGGACTGTTTCCAGAGGATGGTGTGTACCCAGGCTACACACAAAAAGAGAGATAAA 2567
Db 310 IleSerPheLeuGlnGlu-----LysIleLysValTyrGluMetGluGlnAspLys 326
QY 2568 ATATGGAATAATAGAGATTCACTAGCCTATCAAAATCTTGGTACAGTTCATCTGTGA 2627
Db 327 LysValGluAsnSerAsn-----332
QY 2628 AAGGCAAGGAACTTCAAAACAGACACTGTG-----AACAGCTACAGGAAAAATG 2675
Db 333 LysGluGluIleGlnGluLysGluThrIleIleGluGluLeuAsnThr---LysIleIle 351
QY 2676 GAACAATGAAAAAGAGTTTGT-----TGACTGCAAAAAGAACTG 2714
Db 352 GluGluGluLysLysThrLeuGluLysAspLysLeuThrAlaAspLysLeuLeu 371
QY 2715 TCAGACAAAGAAATAATCAATCACTAGAGAACCAAAAGTTAAATGGGAACAAAGACTGTG 2774
Db 372 GlyGluGlnGlnGlnIleValGlnLysAsnGlnGluIle-----LysAsnMet 388
QY 2775 CAGTGTAGATTGACTTAAACCAAGAAAAAGAGAGAGAAATGCCGATATATTAAGAAA 2834
Db 389 LysLeuGluLeuThrAsnSerLysGlnLysGluArgGlnSerSerGluGluIleLysGln 408
QY 2835 AAATTAGGAGATTAGAGAAATCGAGACAGC-----2867
Db 409 LeuMetGlyThrValGluGluLeuGlnLysArgAsnHisLysAspSerGlnPheGluThr 428
QY 2868 -----ATAGAAAGAGTTAGAAAGTGAACACAAA-----CTTGAAAGGCTCTC- 2908
Db 429 AspIleValGlnArgMetGluGlnGluThrGlnArgLysLeuGluGlnLeuArgAlaGlu 448
QY 2908 -----2908
Db 449 LeuAspGluMetTyrGlyGlnGlnIleValGlnMetLysGlnGluLeuIleArgGlnHis 468
QY 2909 -----AGATACAAAGATAAGAGTTGAGGTGTAGAAAGTAATTTGAATAGG 2953
Db 469 MetAlaGlnMetGluGluMetLysThrArgHisLysGlyGluMetGluAsnAlaLeuArg 488
QY 2954 TTCTCAACTCATGAATGAAATATTCTTTCATCGAAATTCAT-----GTT 3001
Db 489 SerTyrSerAsnIleThrValAsnGluAspGlnIleLysLeuMetAsnValAlaIleAsn 508
```


QY 3365 ATCAACATTCATTTTCCTTTGCAACAGCATCTCTCAATGCGAGCAGGTTCTACTTT 3306
Db 130 lleuTrpTyTyTyMetLeuAsnLeuIleAsnPhelIleLeuLeuPheIleLeuLeuTy 150
QY 3305 TTCCTGATGTCACATTCATGCTGTGTGACGAGCAGCCAGTCAGGATGGTATTCA 3246
Db 150 r-----PheMe 152
QY 3245 ATTTCGCTCTAGTTTCCTTTCTGTTCTGTTTTCCTCAATTTAGATGAGCATTTCTC 3186
Db 152 tIleLeuAsnTyTyCysPhePheLeuCys-----AspPheCysPheLe 166
QY 3185 AGCTATCAGAACTTTAGCTGCCACATATG-----AGATCCCTTTTAGTAATGAT 3135
Db 166 uValPheAspGluTrpLeuGlyIleLeuCysLeuPheTyThrLeuLeuIleLeu-P 186
QY 3134 TC-----CTTTTCAGTTTTCAGGTCATCTAAGTTCAGCATCTTTCTTTAAATCTAA 3081
Db 186 heLeuLeuTyTyIleAlaPheLeuIleLeuPheMetGluGlnLeuTyTyIleArgLeu--- 204
QY 3080 GTCCTCAAGATTTATTTTCCTTTCCGCGATTGGTGTAGTCTGGCTATCCAGTTT 3021
Db 205 -----GlyValPheIlePheIleTyMet-----LeuThrPheTyTyIleLeuPheC 220
QY 3020 GCATGGCAATTCCTTTTCAACATGCAATTTTCATGTAAGATAATTTTCATTCATGAGT 2961
Db 220 ysPheIleLeuIleLeuLeuIleSerPheIle-----TyPheTyTyIle 236
QY 2960 TCAGAAACCTATTCAAAATCTTCTACACCTTCAATCTTATCTTGTACTGAGACCT 2901
Db 236 euPheIleLeuLeuLeuPheGlnSerCysThrCysValIleIleLeuAsnSerP 256
QY 2900 TTCAGTGTGTGTTTCACTTCTAATCTTCTATGCTGCTCTCGATCTTCTTCAATCTTCC 2841
Db 256 heAlaIleVal--SerLeuLeuPheValLeuSerValAsnAsnPheCysPheLeuPhe 275
QY 2840 TAAATTTTCTTAAATATATATGCGAATTCCTCTCTCTCTTCTTCTTCTTCTTCTTCT 2781
Db 275 euIlePheIleSerThrLysAsnTyTyIlePheTyTyLeuTyTyLeuAsnPheHisLeu 295
QY 2780 ACACGTCAGACTCTGTTTCCCACTTAACCTTTTGTGTTCTCTACTGATGATTTCTTTG 2721
Db 295 yzSerIle-SerLeuVal-----LeuLeuIleIleIleTyTyTyPhePheIleIle 311
QY 2720 TTCTGACAGTCTTTTTCAGTACACAAACTCTTTTCATTTGTTCCATTTTCTGTACGT 2661
Db 312 TyAsnIlePheAspPheLysTyAsnGluAsnTyTyPheLeuIleAsnPhe----- 328
QY 2660 GTTCACAGTCTTTTTCAGTCTTCTTCCCTTCCCTTCCACAGATGACTCCAAAGATTTT 2601
Db 329 -----IlePhePheSerPhe-----PheAsnAsnPhe 337
QY 2600 GATAGGCTAGTGAATCTTCTA----- 2580
Db 338 -----LeuIleSerLeuLeuLeuAlaCysLeuPheLeuCysIleGlyAlaIleProile 355
QY 2579 ATTTTCCACTATTTTATCTTTCTTTTGTGTG 2547
Db 356 ValPheGlyPhePheIleLysValPheCysLeu 366

RESULT 5
S73852
hypothetical protein MG218 homolog F10_orf1818 - Mycoplasma pneumoniae (strain ATCC 2934
A:Species: Mycoplasma pneumoniae
C:Accession: S73852
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105985; PMID:8948633
A:Accession: S73852

A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1818 <HIM>
A:Cross-references: EMBL:AE000051; GB:U00089; NID:G1674211; PID:AA896174.1; PID:G1674222
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3
C:Superfamily: Mycoplasma genitalium hypothetical protein MG218

Alignment Scores:
Pred. No.: 0.000125 Length: 1818
Score: 139.00 Matches: 236
Percent Similarity: 35.48% Conservative: 200
Best Local Similarity: 19.20% Mismatches: 384
Query Match: 1.96% Indels: 409
DB: Gaps: 56

US-09-602-362E-22 (1-4115) x S73852 (1-1818)

QY 901 AAGTTGGAAGCTCAGCAGAAACACCTAGGAAATTCAGTCTCTCAGAAACAAACATCTGA 960
Db 506 LysLysGlnLysGlnAspGlnLysGluAsnAspLeuLeuIlePheGluLysGlnLeuArg 525
QY 961 GAAATTTACGGCCAGCAAGGAGACCTAGAGATCGCAGGAGAAAGAAAGACACAC 1020
Db 526 GlnTyT-----GlnAlaAspPheGluAsnGluIleGluGluLysGlnAsnGlu 541
QY 1021 CTAGGGAATATAGTCTCCGAAAGAAACACTAGAAATTCAGTGGGAGAAAGAAAGA 1080
Db 542 Leu-----PheAlaSerGlnLysSerLeuGlnLysSerPheThrGlnLysAsnLys 559
QY 1081 CCTAGGAGACCCATCGGAGAGAAAGAAACCTGTAAAGATGGAT-----GCGTGCAAG 1134
Db 560 GluAlaGluLeuAsn---GlnLysAlaGlnLysIleAlaGluAspTrpAlaHisLeuLys 578
QY 1135 AGTAACATCTAATAAACTAAAGTTTGAAGAGAAATCTAAGATGTTGCATGT----- 1188
Db 579 GlnAsn-----LysHisHisAlaAspLeuLeuIlePheLeu 591
QY 1189 -----CCACAAAGAAATCATCTCAAGAGCAAGTCCCATGATCAGAGTCCCATCAGATCAA 1245
Db 592 GluGlyGluPheAsnHisLeuGlnGlnLysHis-----LysLeuLeuGluAlaArg 609
QY 1246 ACAAGGGAAGATGAAGAAATATTTTGTATTTCTCGGGTCTCTTTGAAGTTCGCAAGATTC 1305
Db 610 ThrGlnPheAspAsnArgValSerLeuLeu-----SerAlaArgPhe 623
QY 1306 AAGTGTATACCTGAGTCTATATATAAAAGTAATGAGATAAATAAGAGTAGAAAGCCT 1365
Db 623 ----- 623
QY 1366 CCTAAAGCCATCTGCCTTCAGCCGCTTGAAGCAAACTCTTTCCAAATAAGCCTT 1425
Db 624 -----LysGlnLysGlnAlaGluLeu----- 630
QY 1426 TGAATGAAGAATGAACAAACATTTGAAGCAGATCCGCTGTTCACCCAGATCCAAAGG 1485
Db 631 ---ValLysGlnLysGlnSerLeuGluGlnLeuThrAlaAlaPheAsnLysGluGlnGlu 649
QY 1486 ACTATAAGAAATTTCTGGATCTTAGAGTCTCTGGAGCTCTTTTACAGAG----- 1537
Db 650 AlaVal-GluArgAspTrpLysAspArgLeuAlaAsnLeuGluLysGlnLysGluMetLe 669
QY 1538 ---GATTGTGTTTCCCAA---GGCTACACATCAAAAGAAAGATAGATAAAATCGAAAA 1590
Db 669 uGlyAspLysValHisGlnPheAspGluAsnSerLeuAsnIleSerLysLysLeuAlaG 689
QY 1591 TAGAGAGTCCCTAATAAGTGG----- 1613
Db 689 uArg-GluLeuAlaIleLysPheLysGluLysGluLeuGluAlaGlnLysGlnLeu 709
QY 1614 -----TCTTCTGAAGGCTACCTCGCAATCAAAATTT---CTATTCCCACTAAAGCCTAGAAT 1665
DB: |||

Db 709 erLeuAspAsnAsnAsnAlaGlyLeuLysLeuGlnLeuAspLysLeuSerGluSerL 729
QY 1666 TGAAGACATGCAAACTTTCAAAGCGAGCTCCGGGAAG----- 1703
Db 729 euLysThrGluArgLeuGluLeuAlaSerLysGluArgIleLeuAspPheTyrAspG 749
QY 1704 -----CCATCTGCTT 1713
Db 749 luSerSerArgArgIleAlaAspTyrGluSerAspLeuGlnAlaArgLeuAlaGluValL 769
QY 1714 CGAGCTCCACTGAAATCAAAAGTCTGTCCCAAAATAAG----- 1751
Db 769 ysThrLeuGluLysAsnGlnGlnThrAlaAlaLysSerGluArgGluLeuLysValA 789
QY 1752 -----CCTTGAATGAAATGAAACATGAGCAGCAGATGAGAACTCCCATCAGAA 1803
Db 789 laLeuGluLysLeuAsnGlnAlaLysLysAlaPheLeuGlnIleArgLysGlnLeuL 809
QY 1804 TCCAAACAAAGCACTATAAGAAATTCCTGGGATACTCGAGTCTCTGTAGACTGTTCCACA 1863
Db 809 euGluIleAlaSerValLysGlnLeuAla----- 819
QY 1864 GAAGATTTGTTTACCAAGCTCGCTCAAAAGAAATAGATAAAATAATGGAATAATAG 1923
Db 820 -----GlnLysAlaAsnLeuLeuLysAsnGln-----G 829
QY 1924 AAGGTCCTCTTAAAGTGTCTTCTGAGGCTAACTCGGAATGAAAGTCTTATTCCACT 1983
Db 829 InAlaGluLeuAspLysGlnThrGlu----- 837
QY 1984 AAAGCCTAGAAATGATGACATGCAAACTTCAAAGCAGCCTCCGAGAGCATCTGCCTT 2043
Db 838 -----GluLeuGluAlaAlaPheLeuGluGlnAspThrAspLysLysGluLeuGluL 855
QY 2044 CGAGCTCCATTTGAATCAAAAGTCTGTCCAAATAAACCTTGAATTAAGAAAGCAAA- 2102
Db 855 ysAlaLeuHis-----SerValLysSerLysGlnG 865
QY 2103 --ACATTGAGCAGATGAGAACTCCCATCAATCCAAACAAAGCACTATGAGAAAGTTCT 2160
Db 865 luLeuLeuGluArgGluArgSerPheLeuLeuGlnLysGlnArgGluPheAlaGluHisV 895
QY 2161 TG-----GGATTCGAGTCTGTGTAGACTGTTTCCAGAAAGATGTGTTCACCGAGGTAC 2217
Db 885 alAlaGlyPheLysArg-GlnValHis-----Phe 894
QY 2218 ACATCAAAAGAAATAGATAAAATAATGGAATAATAGAGTCTCTGATAATAGGTTTC 2277
Db 895 LysThrThrGlnMetGlnArgLeuSerGluPheAsnLysGlnGlnSerGluGlnIle 914
QY 2278 TGAAGGCTCCCTGAGAATGAAAGTCTTATTC----- 2308
Db 915 LysArgGluThrGlu-LeuLysIleAlaPheAlaAspLeuLysLysAspTyrGlnLeuPh 934
QY 2309 -----CACTAAAGCCTTGAATT 2325
Db 934 eGluLeuGlnLysAsnGlnPheGlnIleGluGlnLysHisLysGluLeuGluLe 954
QY 2326 GATGGCATGCAAACTTTCA-----AAGCAGGCTCCCGAGAGCCATCTGCTTCGAGCCT 2379
Db 954 uLeuAlaGlnLysGlnAlaGluLeuLysGlnGluGluGlnLysAlaThrAlaLeuAl 974
QY 2380 GCATTGAAATGAAAGTCTGTTCAAATAAACCTTGAATTAAGAAATGAAACATTTAG 2439
Db 974 aserGlnAspGlnAspThrVal-GlnAlaLysLeuAspLeuAlaArgGlnGlnHis----- 992
QY 2440 ACAGATCAGATTTCCCTTCAGAAATCAAAACAAAGAGTGTGAGAAATTCGGGATTCG 2499
Db 993 -----GluLeuGluLeuArgGlnAsnAlaPheAsnGlnAlaAs 1005
QY 2500 AAGTCTCCGTGGACTGTTTCAGAGAGGATGGTGTGTACCCAGCTACACACAAAGAA 2559
Db 1005 erLeuSerLeuAsn-----LysGlnA 1012

QY 2560 AGGATAAAATAGTGGAAAAATTAGAGATTCACTAGCCTATCAAAATCTTTGGTACATTCA 2619
Db 1012 rgGlnGlnLeuThrAsnGln-----ValLysValLeu----- 1022
QY 2620 TCTTGTGAAGCAAGGGAACCTTCAAAAAGACACTGTGAACACGTACGAGAAAATGGAAC 2679
Db 1023 -----HisGlyGluLeuLysLys-----ArgHisGluLysLeuThrL 1035
QY 2680 AATGAAAAAGAGTTTTGTGTACTGAAAAAGAACTGTCAAGACAAAAGAAATAAATCACAC 2739
Db 1035 euLysAspArgLeuLeu-----AlaGluLysGluLysAspGlnHisLysLysAspAlaGluI 1054
QY 2740 TAGAGAACCAAAAGATTAAATGGGAACACAGACTCGAGTGTAGATTGACTTTAAACCAAGA 2799
Db 1054 le-----AsnGlnArgPheLysGlnPheGluAsnGluTyrAlaAspPheAspGlnAlaLysL 1073
QY 2800 AAAGA-----GAAGAGAGAAATCGCGATA 2823
Db 1073 ysArgGluLeuGlnGluLeuAsnGlnIleArgArgAsnLeuGlnSerAsnAlaSerL 1093
QY 2824 TATTAAGAAAAA-----TTAGGAAGAATTAGAAAGTCTCGAGAGCA 2865
Db 1093 euLeuLysLysArgAsnGlnLeuThrLeuAspPheAlaLeuLeuArgLysValGlnHisA 1113
QY 2866 GCATAGAA-----AGAGTTAGAGTGAACACAACTTGAAGGCTCTCAGATACAGAT 2919
Db 1113 snThrGlnThrAsnArgValGlnLeuAsnThrGln-IleLysGluPheLeuLeuGluLys 1132
QY 2920 AAGAATTGAAGGTGTAGAAAGTAATTGAATAGTGTCT----- 2958
Db 1133 LysAsn-----PheGlnLysAlaSerAspGluAlaAlaLeuGlnLys 1146
QY 2959 -----CAACTCATGAAATGAAAA 2976
Db 1147 AlaLeuLeuIleLysArgLeuArgSerPheAlaSerLysLeuGlnLeuGlnArgGluAla 1166
QY 2977 TTATTCCTTACATGAAATTCATGTTGAAAAAGGAATTGCGATCGCAAACTGCAATAGCCA 3036
Db 1167 LeuAlaIleGlnLysLeuGluPheAspLysArgAspGluGlnGlnLys----- 1182
QY 3037 CACTAAACCAATCCAGGAAAGGAATAAATCTTTGAGGACTTAAGATTTTAAAGAA 3096
Db 1183 -----SerGluIleAsnAsnAlaLysLeuGlnLeuGluGlnPheLysLeu 1197
QY 3097 AAGATGCTGAATAGATGACCTAAACTGAAAGGAATCAATTAATAAGGATCTCA 3156
Db 1198 GluLysGlnAsnPheAspGluAlaLys-----GlnLysGlnLeuIleGluPheLysAsp 1215
QY 3157 ATATGTGGCAGCTA-----AAGTTCTGTAGTCTGAGA 3189
Db 1216 GlnCysGlnArgLeuAspValGluLysArgLeuLeuLysGlnLysLeuValGlnLeuLys 1235
QY 3190 AACATGTCTCATCTTAATTTAGG-----RAAAACACAGACAAAGAAACTAGAGCGCAAAAT 3246
Db 1236 AsnLeuSerLysSerTyrLeuThrTyrLysAsnArg-AlaAspLeuSerGlnGlnLeu 1255
QY 3247 GAATACACCATCTGACTGGCTTCGCTGTACAGCCATCATCAATTGTGCATCAAGAAA 3306
Db 1255 u-----GlnHisly 1258
QY 3307 AAGTAAGAACCTGCTTCCACATTTAGGAGATGCTGTTTGCAGAAAGAAATGAATGTTGATG 3366
Db 1258 styAlaAsnLeuLeu-GluLeuLysGluLysLeuGlnThrAlaLysArgAlaLeuAspL 1278
QY 3367 TCAGAGTAGATTAATCAACAATGAGTCTCCAT-----CACCACTTTCTGACTCAA 3417
Db 1278 ysLysHisArg-----AlaIleTyrGlyLysMetAlaGlnPheValSerGluL 1294
QY 3418 AGGAATCCAAACCTCAAAATATCTCAAT-----ACCAAGA 3456
Db 1294 euArgGlnGluLysLysGlnLeuSerAlaGlnLysGlnValAspAspLysSerArgL 1314


```
QY 3099 CTTTTCCTTAA-----ATCTTAAGTCTCAAGATTATTTCTTTTC 3055
    |||||
Db 371 TyrPheThrTyrPheCysGlyCysileLeuAlaArgMetLeuLeuilePheAlaile 390
QY 3054 CTGGATTGGCTTTAGTGTGGCTATTTCAGTTTTCATGGCAATTCCTTTTCAACATGC 2995
    |||||
Db 391 LeuAsnSerCysSerLeuTrpPheLeu-----Cys 400
QY 2994 AATTTCATGTAAGAATAATTTTCATTTTCATGAGTTGAGAACTTATCAAAATTAATTCT 2935
    |||||
Db 401 GlyPheTyr-----CysLysAspLeuLeu----- 408
QY 2934 ACACCTCAATCTTATCTGATCTGAGACCTTTTCAAGTTGTGTTCTACTTCACTC 2875
    |||||
Db 409 -----LeuCysLeuLeuMetLeuThrSerPhePheileLeuGluPhe-----Leu 424
QY 2874 TTCTATCTCTCTCGATCTTCTTAATCTCTCTCAATTTTCTTAAATATATATCGGCAT 2815
    |||||
Db 425 CysValCysLeu-----PhePheilePhePheThrValleTyrAsnTyr 439
QY 2814 TTCTCTCTCTCTTTCTTTG-----TTTAAGTCAATCTACACTGCGAGACTCTTTGTCC 2761
    |||||
Db 440 PheLeuLeuPhePheLeuCysPheValPheLysCysPheCysLeuValAspThrLeuPhe 459
QY 2760 CATTTA---ACTTTTGGTCTCTACTGATGATTTATTTCTTTTGTCTGACAGTTCTTTT 2704
    |||||
Db 460 LeuLeuPheAspPheGluCysCysLeuValTyrCysThrPheCysLeuTyrMetCysPhe 479
QY 2703 CAGTACACAAACCTTTTTCATGTTGCCATTTTCTCTGTAGC-----TGTTTCACAG 2653
    |||||
Db 480 -----ValLeuLeuPhePheValLeuAspPheLeuTyrValPheilePheSerSerTyr 497
QY 2652 TGTCTTTT 2644
Db 498 CysLeuPhe 500

RESULT 7
CQUOTCS
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Trypanosoma brucei mitochondrion
C:Species: mitochondrion Trypanosoma brucei
C:Date: 17-Mar-1987 #sequence_revision 17-Mar-1987 #text_change 03-Jun-2002
C:Accession: A04519; D28845
R:Hensgens, L.A.M.; Brakenhoff, J.; De Vries, B.F.; Sloof, P.; Tromp, M.C.; Van Boom, J.
Nucleic Acids Res. 12, 7327-7344, 1984
A:Title: The sequence of the gene for cytochrome c oxidase subunit I, a frameshift contain-
ondrial maxi-circle DNA
A:Reference number: A93537; MUID:85037915; PMID:6093040
A:Accession: A04519
A:Molecule type: DNA
A:Residues: 1-590 <HEN>
A:Cross-references: GB:M94286; NID:g343546
A:Note: this translation is not annotated in GenBank entry TRBKSGEN, release 109.0
C:Comment: The DNA sequence is from a segment of the 20-kb maxicircle, which is believed
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC6
A:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 0.000126 Length: 590
Score: 138.00 Matches: 107
Percent Similarity: 34.08% Conservative: 75
Best Local Similarity: 20.04% Mismatches: 162
Query Match: 1.92% Indels: 190
DB: 1 Gaps: 26

US-09-602-362E-22 (1-4115) x QQUOTCS (1-590)
QY 3983 ACTTTATTTATGCTGAGCTTCAGGAAGAAACATGAATTTCTTTCTAAATAGTATTC 3924
    |||||
Db 194 ThrThrPheTyrPheLeuAsnPhePheCysMetAspTyrTyrTyrileGluPheSerile 213
```

```
QY 3923 TGT-----GACACAGACTTTGGTAAGATTTTCAACATA 3891
    |||||
Db 214 CysLeuLeuValGlyCysAlaPheThrLysSerThrGlnPheGlyLeuHisileTrpLeu 233
QY 3890 GGTGATGCTAGACTGCGCTCTAGCTAGCTAGTGAAGTAAAGTCTGCTCTGTTTCTTCCCA 3831
    |||||
Db 234 ProAspAla---MetGluGlyProileProValSer-----AlaLeuileHisAla 249
QY 3830 AAGAAGTTCTACTG-----CTTGTCTCATGAGTTTC-----TGTTTCTCTTTC 3786
    |||||
Db 250 AlaThrLeuValValCysGlyileleleLeuLeuSerPheValTyrTrpCysPheAspPhe 269
QY 3785 TCTTTTCA-----TATTGAATATACGTTTAAATGGTATATGTAATAATATCTCT 3732
    |||||
Db 270 TrpPheSerTyrPheTyrAsnLeuileGlyTrpSerThrLeuileLeuileLeuMetThr 289
QY 3731 CTTTTTCTCTTTTAGGAGATGTTGCAATTCCTCTCAAGAAATGAATATCATTTGTTATC 3672
    |||||
Db 290 LeuCysValPhe-----TyrAsnPheAspVallyAsrGlyrValAlaPhe 304
QY 3671 TGTCTTTTGTGTCAGCTTTCTATGTGCATGACTAATTTGCTGTGAAGCCACAAATTTTGTCT 3612
    |||||
Db 305 SerThrileCysGlnileSerPheSerMetPheCysCysLeuCysileAspileTyrile 324
QY 3611 TGTAGTTGAATAATTTCTGACTAGAGACTCTGCTGTTTCAGGTGT----- 3567
    |||||
Db 325 GlySer-----LeuPhePheCysTyrHisMetPheTyr 335
QY 3566 -----TTGTTTCATTATCTTGTGCTTTTGATACATGTTCTGCTTCTCTTCATTGAC 3516
    |||||
Db 336 LysAlaThrLeuPheileValleuGlyileTrpilleHisilePhePheGlyLeuGlnAsp 355
QY 3515 ACTGTTTTACGTTGTCTCTTTGTGATGTTCTGAACCAATATTTCTCTTAGACATCT 3456
    |||||
Db 356 -Leu-----ArgCysTyrPhePheMetTyrPheCysGlyCysValLeuAlaAargLeu 373
QY 3455 CTGCTAATTGAGATAAATTTTAGGTTTGGATTTCTTTGAGCTCAGAAAGTGGTGATG 3396
    |||||
Db 373 uleuillePheAlaileLeu----- 379
QY 3395 GAGCACTAATGTTAATATCGTACTCTCACATCAATTCATTTCTTTGCAACAGCAT 3336
    |||||
Db 380 -----AsnSerCys 382
QY 3335 CTCTCAATGTGGAGCAGGTTCTTACTTTTCTGATGTCACAATGTGATGCTGCTGT 3276
    |||||
Db 382 sSer----- 383
QY 3275 ACAGCGAGCCAGTCAGGATGGTGTATTCAATTTCTGCTCTAGTTTCTTTGCTGTTT 3216
    |||||
Db 384 -----ileTrpPheLeuCys-- 388
QY 3215 TTCCTCAATTTAGATGACCAATTTCTCAGCTATCAGAACTTTAGTCGCCACATATT 3156
    |||||
Db 389 -----GlyPheTyrCys-----LysAspMetle 396
QY 3155 GAGATCCCTTTTAGTAATGATTCCTTTTTCAGTTTGTAGGTCATCTAAGTACGATCTTT 3096
    |||||
Db 396 uleuAlaLeuLeuMetLeuLeuSerPheTyr-----Asnille-ileGluPheLeu 413
QY 3095 TCTTTAAATCTTAAGTCTCAAGATTTATTTTCTCTT-----TTC 3054
    |||||
Db 413 heileSerileile-----PheilePhePheThrMetileTyrAsnTyrPheLe 429
QY 3053 TGGATTGGTGTGTAGTGGCTATTTCCAG-----TTTGTGATGGCAATTCCTTTTCAACAT 2997
    |||||
Db 429 euLeuPhePheLeuMetPheValPheLysCysPheCysLeuVal-----AspC 445
QY 2996 GCAATTTCATGTAAGAATAATTTTTCATTTTCATGAGTTGAGAAACCTTATTAATTTACTTT 2937
    |||||
Db 445 ys-----LeuPheLeuLeuPheA 451
QY 2936 CTACACCTTCAATTTCTTATCTGTATCTGAGAGCCTTTCAAGTTGTGTTTCACTTCTAAC 2877
```


QY 2544 GCCTGGTACACACCATCTTCTGTGAAGACAGTCCACGAGACTTCAGATCCCGAGATT 2485
Db 434 hePheGly-----ProLeuVal-----P 440
QY 2484 TTCTCAACCTCTTTGTTTGTATTTCTGAAGGAAATCTGATCTGTCTCAATGTTGTTTCAT 2425
Db 440 heLeuPheMetMetilePhe-----LeuSerP 449
QY 2424 TCTTAATTCGAAGGTTTATTTGGACACAGACTTTTCATTTCAATGCGAGCTCGAAGCAGAT 2365
Db 449 heLeuileLeuLysMetLeuPheLysGluLeuMetTyrLysPheLeuVal-----AspT 467
QY 2364 GCCTCTCGGAGGCTCTTTGAAAGTTTGCATGCCATCAATTCAGAGCTTTAGTGGAT 2305
Db 467 yrLeuAlaLysAsnSerileTyrLysMetLysAsnLeuLysPheMetAspLeuLeuA 487
QY 2304 AGAATTCATCTTCAGGGA-----GCCTTCAGAAAACCTCAT 2266
Db 487 snAsnileAsnSerLysGlyTyrThrLeuPheLeuSerSerGlyMetPheLysAsnTyrT 507
QY 2265 ATCAGAGACTCTCTATTTT-----CCATTATTTATCTATTCTTTTGTGATG 2218
Db 507 yrLeuLysSerLeuAsnPheAsnSerValValLeuilePheilePheMetileC 527
QY 2217 GT 2216
Db 527 ys 527
RESULT 12
T44522
hypothetical protein 4S [imported] - Shigella sonnei
C:Species: Shigella sonnei
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T44522
R:Chida T.; Okamura, N.; Yoshida, Y.; Ohtani, K.; Arakawa, B.; Watanabe, H.
submitted to the EMBL Data Library, May 1999
A:Description: Complete DNA sequence of the O-antigen (rfb) gene cluster in Plesiomonas
A:Reference number: 222787
A:Accession: T44522
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-426 <CHI>
A:Cross-references: EMBL:AB028135; PIN:BA05073.1
A:Experimental source: isolate HW383
Alignment Scores:
Pred. No.: 0.000367 Length: 426
Score: 133.00 Matches: 108
Percent Similarity: 37.42% Conservative: 78
Best Local Similarity: 21.73% Mismatches: 169
Query Match: 1.85% Indels: 143
DB: 2 Gaps: 21
US-09-602-362E-22 (1-4115) x T44522 (1-426)
QY 3856 GTAAAGTCTGGTCTGTTGTTCTCCAAAGAGTTTCTACTGCTTG-----TCGCA 2816
Db 6 ileAspAlaGlyGlyThrPheLeuLeuLysAlailePheGlnileGlyValPheValTyr 25
QY 3811 TTCTCATGAGTTTCT-----GTTCTCTTCTCTCTTTTTCATATTGAATATAGCGTTT 3758
Db 26 PheThrHisValSerAspilleThrPheGlyileleSerTyrValPheThrValTyr 45
QY 3757 AATGGATTGTAATTAATATCTCTCTTTTCTTTTATAGGAGATGTGTGCAATTCCT 3698
Db 46 ---TrpPhe-ValLeuAsnPheSerAspTyrGlyPheArgThrLysLeuValLysAsp11 64
QY 3697 CTCAGAAAGTAATATCATTTGTTATCTGCTTTGTTGTCAGCTTCTATGTGTCATGACTA 3638
Db 64 eSerAspAsnSerTyrSerAlaSerGluLeu----- 74
QY 3637 ATTGCTGTGAAGCACCAATTTTGTGTGTAGTTGAAATAATTTCTGACTAGACTCTGC 3578

75 -----LeuSerArgSerAspG1 80
3577 TGTTCAGGTGTTGTTCCATTATCTTGTGCTTTTGTATACATGTTGCTTCCTTCATG 3518
Db 80 yValLysThrTyrVal-PhePhePhelellePheMetPheTyrSerTyrValis 100
QY 3517 ACATGTTTTCAGCTGCTCTTGTGATGTTCTCGAAACCA-----TGTATT 3470
Db 100 exAspSerileSerLeuThrLeu-LeuValTyrileSerSerAlaTyrPheValCysile 119
QY 3469 TCTCTTAGACATCTCTGCTTAATGAGATAATTTTGTGTTTGGATTTCCTTTGAGCTC 3410
Db 120 SerSerGlyArgPheSerLeuLeuGlnAlaValGlyArgPhe-----ArgCysGluLeu 137
QY 3409 AGAAAGTGTGATGGAGCACTCATTTTAATATCTGTACTCTCACATCAACATTCATTTTC 3350
Db 138 -----TyrileAsnileTyrSerThrilleleTyr 147
QY 3349 TTTGCAACAGCATCTCCTCAATGTGGAAGAGCTTCTACTTTTCTTGATGTCACAT 3290
Db 148 -Ile-----GlyCysAsn-----LeuPheLeuSerLeuPhe11 158
QY 3289 TGATCATGCTGTACAGCGAAGCCAGTCAGATGCTGTATTCAATTTCTGCTCC----- 3237
Db 158 eGluProLeuTyrTyrSerAlaileSer-IlePheileTyrSerileSerLeuLeuValP 178
QY 3236 --TCTAGTTTCTTGTCTGTTTCTCTCAATTTAGATGAGCATTTCTCAGCTATC 3179
Db 178 heSerSerHisLysCys----- 183
QY 3178 AGAACTTTAGCTGCCACATATTAGATCCCT-----TTTA 3143
Db 184 --AsnValProCysPheHisileLysArgProSerileLeuValTyrLysAspPheLeuA 203
QY 3142 GTAATGATTCCTTTTTCAGTTTGTAGGTCAATCAAGTTCAGCATCTTTCTTTAAATCTT 3083
Db 203 spAlaThrProPheAlaileLeuValLeuLeuAsnValValLeuSerSerileAspLeuP 223
QY 3082 AAGTCTCTCAAGATTATTTTCTCTTCTGATGCTGTTTGTAGTGTGCTATTCAGTT 3023
Db 223 heileLeuLysGluTyrPheSer-----TyrAsnSerV 234
QY 3022 TTGCATGGCAATTCCTTTTCAACATGCAATTTCAAGAAATAATTTTCATTCATCA 2963
Db 234 alAla-----ileTyrGlnValValThrArgValAsnThrGlyLeuileleValP 251
QY 2962 GTTGAAACCTATTCAAAATTAATTTCTACACTTCAATCTTATCTTATCTGAGAGC 2903
Db 251 heAsnValileTyrThr-ValLeuLeuProSerPheSerTyrTyrLeuLysAsnSerGlu 270
QY 2902 CTTTCAAGTGTGTTTCACTTCTTCTATGCTGCTCTGATTTCTTCTTAATTC 2843
Db 271 TrpGlyAsnileArgLysLeuGlnArgTyrileSerLeuLeuValLeuLeuCysLeu 290
QY 2842 CTAATTTTCTTTAAATATA-----TCGCA 2816
Db 291 CysTyrTyrPhePheGlyileTyrPheValGlyileLeuPheGlyAspGluTyrLysVal 310
QY 2815 TTCTCTCTCTCTCTTTTCTGTTTAAAGTCAATCTACACTGCAGACTCTGTTCCCATTT 2756
Db 311 leSerSerAlaThrPheleulle-----MetPheMetAlaLeuileLysTyr 326
QY 2755 AACTTTTGTCTCTACTGCTGATTTTCTTTTCTGACAGCTTCTTTTTCAGTACAC 2696
Db 327 AsnPheTyrPheleulleAsnGluLeuTyrLeuValCysSerGlyAsnGlnSerGluArgVal 346
QY 2695 AAAACT-----CTTTTTCATTTGTTCCATTTTCTGCTAGCTGTTCACA--- 2654
Db 347 LysSerTyrCysileGlyValileSerMetAlaValPhePheTyr-PheileProArg 366
QY 2653 -----GTGCTTTTGAAGTTCCCTTGC-----TTT 2627

Db 367 TyrGlyTrpSerGlyAlaValPheGlySerAlaIleAlaThrLeuValIleGlyIlePhe 386
 Qy 2626 CAC-----AGATGAAGTGTACCAAGATTTTGAT-----AGG 2595
 Db 387 TyrIleSerValLysAspCysGlyLysIleLeuHisAspLysTyrSerLeuMet 406
 Qy 2594 CTAGTGAATCTCTAATTTCCACTATTATCTCTCT 2556
 Db 407 MetIlePheValProllePhePheTyrPheIleIleAsn 419

RESULT 13
 C71622
 Hypothetical protein PF0145C - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: C71622
 R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 A:Accession: C71622
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1979 <GAR>
 A:Cross-References: GB:AE001375; GB:AE001362; NID:G3845105; PIDN:AACT71819.1; PID:G384510
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PF0145C

Alignment Scores:
 Pred. No.: 0.000618 Length: 1979
 Score: 132.00 Matches: 119
 Percent Similarity: 39.22% Conservative: 101
 Best Local Similarity: 21.21% Mismatches: 193
 Query Match: 1.86% Indels: 149
 DB: 2 Gaps: 29

US-09-602-362E-22 (1-4115) x C71622 (1-1979)

Qy 2496 TCTGAAGTCTCCGTGAGTCTTTTACA-----GAAGGATGGTGTGTACCCAGGCTACAC 2549
 Db 1111 SerLysIleAspIleAsnSerLeuThrIleAsnGluGlyMetCysVal----- 1126
 Qy 2550 ACAAAGAAAGGATAAATAGTGGAAATTAAGAATTCAGTCTAGCTATCAAAATCTTG 2609
 Db 1127 -----AspLysIleGluGluHis----- 1132
 Qy 2610 GTACAGTTCATCTGTGAAAGGCAAGGAACTTCAAAAGACAGCTGTGAACACGTACAGG 2669
 Db 1133 -----IleLeuAspTyrAspGluGluIleAsnLysSerArgSerAsnLeuPhe--- 1148
 Qy 2670 AAATGGAACAATGAAAGAGATTTTGCT-----ACTGAA-----AAGAAGCTGCA 2717
 Db 1149 -----GlnLeuLysAsnGluIleCysSerLeuThrThrGluValMetGluLeuAsn 1165
 Qy 2718 GAACAAAGAAATAATACACTAGACCAACCAAAAGTTAAAT-----GGGAAC 2765
 Db 1166 AsnLysLysAsnGluIleGluAsnAsnLysLeuAsnLeuValAspGlnGlyLys 1185
 Qy 2766 AAGAGTCTCAGTGTAGATTGATTTAAACCAAGAAAGAGAGAGAGAAATGCCGATATA 2825
 Db 1186 LysLysLeuLysLysAspValGluLysGlnLysLysGluIleGluLysLeuAsnLysGln 1205
 Qy 2826 TTAAGAAAAAATAGGAAGATTAAGAATTCGAGAGCAGATAGAGAGATTAAGAAGT 2885
 Db 1206 LeuThrLysCysAsnLysGlnIleAspGluLeuAsnGluValGluLysLeuAsnAsn 1225
 Qy 2886 GAAACACAATTTGAAGGGCTCTCATGATACAAGATA-----AGAATTGAAGGTGTA 2935
 Db 1226 Glu---AsnIleGluIleThrTyrSerAsnAspLeuAsnAsnLysPheAspMetLys 1244
 Qy 2936 GAAAGTAATTTGAATAGGTTCTCACTCATGAAATGAAATTTATCTTACATGAATTG 2995

Db 1245 GluAsnAsnLeuMetMetLysLeuAsnGluAsnGluAsnIleLysLysMetLysSer 1264
 Qy 2996 CATGTTGAAAAGGAATTTGCCATGCAAAACTGGAATAGCCACTAAACACCAATCCAGG 3055
 Db 1265 LysIleAsp-----AspMetGluLysGluIleLysTyrArgGluAsp 1278
 Qy 3056 AAAAGCAAAATAAATCTTTGAGCACTTAAGATTAAAGAAAGATGCTGAACCTTAGATG 3115
 Db 1279 GluLysLysArgAsnLeuAsnGluIleAsnAsnLeuLysLysLysAsn---GluAspMet 1297
 Qy 3116 ACCCTAAACTG---AAAAGGAATCAATTAATAAAGGATCTCAATATGTGGCAGCTAA 3172
 Db 1298 CysIleLysTyrAsnGluMetAsn-IleLysTyrGlyAspIleCysValLysTyrGluGln 1317
 Qy 3173 AGTTCTGATAGCTGAGAAACAATGCTCAATCTTAATGAGAAACAAGACAAAGAAAAC 3232
 Db 1317 uMetSerLeuThrTyrLysGluThrSerLeuLysTyrGluGlnIleLysValLysTyrAs 1337
 Qy 3233 TAGAGCAGAAATGAATACACCATCTGACTGGCTTCGCTGTACAGCCCATGATCAATT 3292
 Db 1337 pGlu-----LysCysSerGlnTyrAspGluIle 1346
 Qy 3293 GTGACATCAA-----GAAAAGTAAGAACCTGCTTCCACATTTAGGAGATG-----CT 3340
 Db 1346 eargPheGlnTyrAspGluLys-----CysPheGlnTyrAspGluIleAsnLysLys 1363
 Qy 3341 GTTTGCAAAAGAAATGAATGTTGATGTGAGAGTACGATATTAACATGAGTCTCCATCA 3400
 Db 1363 sTyrGlyAlaLeuLeuAsnIleAsnIleThr-----As 1374
 Qy 3401 CCACTTTCTGAGCTCAAG-----GAAATCCAAACCTTAAATAATTCTCAATTAGCAGAG 3457
 Db 1374 nLysMetValAspSerLysValAspArgAsnAsnAsnGluIleIleSerValAspAsnLys 1394
 Qy 3458 ATGCTCTAGAGAAATACATTTGTTTCAGAACATCACAAGACAGACACCTGAA----- 3509
 Db 1394 sValGluGlyIleAlaAsnTyrLeuLysGlnIlePheGluLeuAsnGluGluIleAla 1414
 Qy 3510 -----AACAGTGTCAATGAA----- 3524
 Db 1414 GluLysGlyGluIleAsnLysIleSerLeuLeuTyrSerAsnGluLeuAsnGluLysAs 1434
 Qy 3525 -----GGAGCGCAACACATGTATCAAAACGCAAGATAATGGAACAAACACC 3571
 Db 1434 nSerTyrAspIleAsnMetLysHisIleGlnGluLeuLeuPheLeuGluLysThrAs 1454
 Qy 3572 TGAACAGCAGAGTCTCTAGTCAAAATTTTCAACTACA----- 3611
 Db 1454 nLysGluAsnGlu-----GluLysIleIleAsnLeuThrSerGlnTyrSerAspAlaTy 1472
 Qy 3612 -----AGCAAAAATTTGTGCG----- 3626
 Db 1472 rLysLysLysSerAspGluSerLysLeuCysGlyAlaGlnPheValAspValAsnIle 1492
 Qy 3627 -TTCACAGCAATTAGTCATGCACATAGAAAGCTG----- 3659
 Db 1492 eTyrGlyAsnIleSerAsnAsnIleArgThrAsnGluTyrLysTyrGluGluMetPh 1512
 Qy 3660 -----ACAACAAAAGCAGATAACAATCATATTTCTTTGAGAGGAAATGCAACATCT 3715
 Db 1512 eAspThrAsnIleGluGluLysAsnGlyMetHisLeu-----SerLysTyrIleHisLe 1530
 Qy 3716 CCTAAAGAAAGAAAGAGGATATTAATACATACCATTTTAAACCGTATATCAAT 3775
 Db 1530 uLeuGluGluAsnLysPheArgCys-----MetLysIleIleTyr-GluA 1545
 Qy 3776 ATGAAAAGAGAGAGAAACACAGAACTCATGAGAACA---AGCAGTAGAAACTCTTTTG 3832
 Db 1545 snGluAsnIleLysSerSerAsnLysIleIleGlyLeuTyrAsnTyrSerArgTyrTyG 1565
 Qy 3833 GAGAAACACAGACAGACTTTTACTCAGCTCATGTAGAGCCAGCTCTAGCATCACCTA 3892

Db 1565 lyleuArgGluAsp-----LeuCysLysGluGluIleValProSerLysI 1580

QY 3893 TCTTGAATAATCTTACAAAGCTGTGTACAGAACTACTATTITAGAGAAATTCATGCT 3951

Db 1580 lcglyAsnIleSerAsnLysAsnGluAsnAsnLys-----LysAsnAsnThrCys 1597

RESULT 14

S36954

Cytochrome-c oxidase (EC 1.9.3.1) chain III - Herpetomonas megaseliae mitochondrion

C;Species: mitochondrion Herpetomonas megaseliae

C;Date: 10-Dec-1993 #sequence_revision 12-May-1995 #text_change 07-Dec-1999

C;Accession: S36954

R;Landweber, L.F.; Gilbert, W.

Nature 363, 179-182, 1993

A;Title: RNA editing as a source of genetic variation.

A;Reference number: S36953; MUID:93247644; PMID:8387160

A;Accession: S36954

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-288 <LAN>

C;Genetics:

A;Gene: COIII

A;Genome: mitochondrion

A;Genetic code: SGC6

C;Superfamily: cytochrome-c oxidase chain III

C;Keywords: electron transfer; membrane-associated complex; mitochondrial inner membrane

in

Alignment Scores:

Pred. No.: 0.000478 Length: 288

Score: 131.50 Matches: 72

Percent Similarity: 39.86% Conservative: 40

Best Local Similarity: 25.82% Mismatches: 91

Query Match: 1.83% Indels: 78

DB: 2 Gaps: 13

US-09-602-362E-22 (1-4115) x S36954 (1-288)

QY 3132 CTCTTTCAGTTTACGGTCATC-----TAAGTTCAGCATCTTTCTTTAAATCTTAAGT 3079

Db 1 MetPheLeuPheArgValIlePheValGlyValSerGlyValPheValPheLeuSerLeu 20

QY 3078 CCTCAAGATTTATT----- 3064

Db 21 ProAlaValValIleValTyrTyrValIleCysLeuCysGlyPheMetIleCysCysPhe 40

QY 3063 -----TTCCTTTCCCGAGTGTGTAGTGGCTATTC-----CAGTTTTCG 3019

Db 41 GlySerPheValPheValAspMetCysPheValPhePhePheGlyLeuLeuPheCys 60

QY 3018 ATGGCAATT-----CCTTTTTCACATGC 2995

Db 61 IleLeuLeuLeuLeuCysAspLeuPheValAspPheLeuArgGlyIlePheAspPheCys 80

QY 2994 AATTCATGTAAGAAATAATTTTCATTTTCAGTTGAGAAACCTATTCAATTAATCTTCT 2935

Db 81 AsnPheLeu-----ArgValLeuGlnTyrCysPhe---MetTrpPhe 93

QY 2934 ACACCTCAATCTTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 2875

Db 94 ValPheSerGlnLeuValIlePhePheMetSerPhePheThrValValPheGlyTyrValIle 113

QY 2874 TTCTATGCTGCTCTCGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2821

Db 114 PheLeuCysCysGluPheAlaPheValPheCysLeuPheProMetLeuPheCysCysLeuLeu 133

QY 2820 CGGCAATTCCT 2761

Db 134 ValAspTyrGlyPheValPheTyrTrpPhe-PheMetAspLeuPheAsnLeuLeuIleAs 153

QY 2760 CATTTAACTTTTGGTT-----CCTACTGTGATTTATTTCTT-----TTGTTCTCG 2716

Db 153 nThrPheLeuLeuPheValSerGlyLeuPheCysAsnPhePheLeuPheCysValTrpPh 173

QY 2715 ACAGTCTCTTTTTCAGTACACAAACCTCTTTTTCATTTGTTCCATT-----TTCCTGT 2665

Db 173 eArgPhePheVal-ValCysIlePheValLeuTrpCysGlyIleLeuPheGlyPheLeuP 193

QY 2664 ACGTGTTTCACAGTGTCTTTTGAAGTTCCCTTCCCTTTCACAGATGAACTGTACCAAGA 2605

Db 193 heLeuTrpAsnGlnLeuTrpGluPheAlaLeuLeuPheValThrValSerCys----- 210

QY 2604 TTTTGATAGGCTAGTGAATCTTCTTAATTTTCCACTATTTTATCCTTTCTTTTGTGTGA 2545

Db 211 -----GlyValPheGlySerIleLeuPheValIleA 221

QY 2544 GCCTG-----GGTACACACCATCTTCTGTGTGAACAGTCCACGGAGACTTCAGATCCC 2491

Db 221 splLeuLeuHisPheThrHis-ValLeuLeu----- 230

QY 2490 AGAATTTTCTCAACCTCTTTGTTGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2434

Db 231 GlyValPheLeuLeuPheIleValPheMetArgLeuPheAsnPheLeuCysMetAsp 249

RESULT 15

S36953

Cytochrome-c oxidase (EC 1.9.3.1) chain III - Herpetomonas samuelpessoai mitochondrion

C;Species: mitochondrion Herpetomonas samuelpessoai

C;Date: 10-Dec-1993 #sequence_revision 12-May-1995 #text_change 07-Dec-1999

C;Accession: S36953

R;Landweber, L.F.; Gilbert, W.

Nature 363, 179-182, 1993

A;Title: RNA editing as a source of genetic variation.

A;Reference number: S36953; MUID:93247644; PMID:8387160

A;Accession: S36953

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-288 <LAN>

A;Cross-references: GB:L10852; NID:G3559760; PIDN:AAD09166.1; PID:G3559761

C;Genetics:

A;Gene: COIII

A;Genome: mitochondrion

A;Genetic code: SGC6

C;Superfamily: cytochrome-c oxidase chain III

C;Keywords: electron transfer; membrane-associated complex; mitochondrial inner membrane

in

Alignment Scores:

Pred. No.: 0.000599 Length: 288

Score: 130.50 Matches: 73

Percent Similarity: 34.70% Conservative: 37

Best Local Similarity: 23.03% Mismatches: 94

Query Match: 1.82% Indels: 114

DB: 2 Gaps: 15

US-09-602-362E-22 (1-4115) x S36953 (1-288)

QY 3132 CTCTTTCAGTTTACGGTCATC-----TAAGTTCAGCATCTTTCTTTAAATCTTAAGT 3079

Db 1 MetPheLeuPheArgValIlePheValGlyValSerGlyValPheValPheLeuSerLeu 20

QY 3078 CCTCAAGATTTATT----- 3064

Db 21 ProAlaValValIleCysTyrTyrValValCysLeuCysGlyPheMetIleCysCysPhe 40

QY 3063 -----TTCCTTTTCCTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 3019

Db 41 GlySerPheLeuPheValAspMetCysPheValPhePheLeuValCysLeuLeuPheCys 60

QY 3018 ATGGCAATTCCTTTTTCACATGAACTTCATGTAAAGAAATAATTTTCATTTTCATGAGTTG 2959

Db 61 IleLeuLeuLeuPhe-----CysAspValPhe-----ValAspPheLeu 73

QY 2958 AGAAACCTATTCAATTAATTTCT----- 2935

Db 74 ArgGlyIlePheAspPheLeuThrPheIleArgCysLeuGlnTyrCysPheIleTrpPhe 93

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: July 15, 2004, 08:50:37 ; Search time 192.589 Seconds

(without alignments)
13356.780 Million cell updates/sec

Title: US-09-602-362E-22

Perfect score: 7086

Sequence: 1 ctactctatcacgaacagc.....gagtggaactccactggaaa 4115

Scoring table: BLOSUM62

Xgapop 10.0	Xgapext 0.5
Ygapop 10.0	Ygapext 0.5
Fgapop 6.0	Fgapext 7.0
Delop 6.0	Delext 7.0

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 2570690

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool.p/US09602362/runat_15072004_093627_22071/app_query.fasta_1.10325
-DB=Published Applications AA -QFWT=fastan -SURFIX=rapb -MINWATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cgi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09602362 -CGEN_1_1 7.19 @runat_15072004_093627_22071
-NCFU=6 -ICFU=3 -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DLEXT=7

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45
1234	17.4	1341	12	US-10-058-270A-4	Sequence 4, Appli																																							
1234	17.4	1341	13	US-10-007-805-565	Sequence 565, App																																							
1234	17.4	1341	14	US-10-076-622-565	Sequence 565, App																																							
1234	17.4	1341	15	US-10-177-293-334	Sequence 334, App																																							
1234	17.4	1341	16	US-10-124-805-565	Sequence 565, App																																							
1223	17.3	1349	13	US-10-007-805-573	Sequence 573, App																																							
1223	17.3	1349	14	US-10-076-622-573	Sequence 573, App																																							
1223	17.3	1349	15	US-10-124-805-573	Sequence 573, App																																							
926.5	13.1	1013	13	US-10-007-805-553	Sequence 553, App																																							
926.5	13.1	1013	14	US-10-076-622-553	Sequence 553, App																																							
917	12.9	1002	9	US-10-124-805-553	Sequence 553, App																																							
917	12.9	1002	10	US-09-604-287A-475	Sequence 475, App																																							
917	12.9	1002	13	US-09-551-621-475	Sequence 475, App																																							
917	12.9	1002	14	US-10-007-805-475	Sequence 475, App																																							
914.5	12.9	1095	14	US-10-076-622-475	Sequence 475, App																																							
914.5	12.9	1095	13	US-10-124-805-475	Sequence 475, App																																							
914.5	12.9	1095	14	US-10-007-805-493	Sequence 493, App																																							
914.5	12.9	1095	13	US-10-076-622-493	Sequence 493, App																																							
612.5	8.6	1239	13	US-10-124-805-493	Sequence 493, App																																							
612.5	8.6	1239	14	US-10-007-805-577	Sequence 577, App																																							
612.5	8.6	1239	13	US-10-076-622-577	Sequence 577, App																																							
597	8.4	661	13	US-10-124-805-577	Sequence 577, App																																							
597	8.4	661	14	US-10-007-805-552	Sequence 552, App																																							
597	8.4	661	13	US-10-076-622-552	Sequence 552, App																																							
597	8.4	661	14	US-10-124-805-552	Sequence 552, App																																							
583.5	8.2	650	9	US-09-825-301-25	Sequence 25, Appl																																							
583.5	8.2	650	9	US-09-604-287A-469	Sequence 469, App																																							
583.5	8.2	650	10	US-09-551-621-469	Sequence 469, App																																							
583.5	8.2	650	13	US-10-007-805-469	Sequence 469, App																																							
583.5	8.2	650	14	US-10-076-622-469	Sequence 469, App																																							
583.5	8.2	650	13	US-10-124-805-469	Sequence 469, App																																							
583.5	8.2	650	14	US-10-033-527-25	Sequence 25, Appl																																							
582	8.2	743	13	US-10-007-805-494	Sequence 494, App																																							
582	8.2	743	14	US-10-076-622-494	Sequence 494, App																																							
544.5	7.7	1225	14	US-10-124-805-494	Sequence 494, App																																							
481	6.8	1011	16	US-10-177-293-332	Sequence 332, App																																							
465	6.6	512	16	US-10-408-765A-1557	Sequence 1557, Ap																																							
421.5	5.9	445	9	US-10-181-663-16	Sequence 16, Appl																																							
421.5	5.9	445	9	US-09-825-301-29	Sequence 29, Appl																																							
421.5	5.9	445	9	US-09-604-287A-473	Sequence 473, App																																							
421.5	5.9	445	10	US-09-551-621-473	Sequence 473, App																																							
421.5	5.9	445	13	US-10-007-805-473	Sequence 473, App																																							
421.5	5.9	445	14	US-10-076-622-473	Sequence 473, App																																							
421.5	5.9	445	14	US-10-124-805-473	Sequence 473, App																																							
421.5	5.9	445	14	US-10-033-527-29	Sequence 29, Appl																																							

ALIGNMENTS

RESULT 1

US-10-058-270A-4

; Sequence 4, Application US/10058270A

; Publication No. US20040029114A1

; GENERAL INFORMATION:

; APPLICANT: Mack, David H.

; APPLICANT: Gish, Kurt C.

; APPLICANT: Afari, Daniel

; APPLICANT: Eos Biotechnology, Inc.

; TITLE OF INVENTION: Methods of Diagnosis of Breast Cancer, Compositions and
; FILE OF INVENTION: Methods of Screening for Modulators of Breast Cancer

; FILE REFERENCE: 018501-005210US

; CURRENT APPLICATION NUMBER: US/10/058,270A

; CURRENT FILING DATE: 2002-01-24

; PRIOR APPLICATION NUMBER: US 60/263,965

; PRIOR FILING DATE: 2001-01-24

; PRIOR APPLICATION NUMBER: US 60/265,928

; PRIOR FILING DATE: 2001-02-02

; PRIOR APPLICATION NUMBER: US 09/829,472

; PRIOR FILING DATE: 2001-04-09

; PRIOR APPLICATION NUMBER: US 60/282,698

; PRIOR FILING DATE: 2001-04-09

; PRIOR APPLICATION NUMBER: US 60/288,590

; PRIOR FILING DATE: 2001-05-04

```

; PRIOR APPLICATION NUMBER: US 60/294,443
; PRIOR FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 141
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-058-270A-4

Alignment Scores:
Pred. No.:      8,89e-110      Length:      1341
Score:          1234.00      Matches:      679
Percent Similarity: 59.18%      Conservatives: 146
Best Local Similarity: 48.71%      Mismatches: 376
Query Match:      17.41%      Indels: 207
DB:              12          Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-058-270A-4 (1-1341)
QY 96 CAAGGGAAGAGACATCAACCTTAATATACAGAG---CCGAGAGAGACTGCTCTAAC 152
DB 3 LysargylsysthrileasleuAsnleuGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGGTCAATGGCT-----GAGGAGTACTAAATTC---GGTA 194
DB 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38
QY 195 GACAGAGTCCAGCTGACCTCTTC---TGGCGAACACGACACCTCTGTATGAGGC-T 250
DB 38 spargylsCysGlnLeuAspValLeuAspValLeuHisArgThrProLeuMetLysAla 58
QY 251 TACAATCCACAGAGGCTTTC---AAATATTGATAGATTC-CGTCCGATATATATC 306
DB 58 euGlnCysHisGlnGluAlaCysAlaAsnleuLeuLeuLeuLeuLeuLeuLeuLeu 78
QY 307 TCCTG---ATGTGTATGCAACATGGCTTCCATTATGCGTTTATAGTGGATTTTGTACGT- 362
DB 78 euValAspValTrpGlyAsnMetAlaLeuHisTrpAlaValTrpSerGluLeuLeuSerV 98
QY 363 --GGTGGCAACTGCTGTC-CATGCTGCAATGCAAGTGC---CAACAAGCTGCTCTCA 416
DB 98 alValAlaLysLeuLeuSerHisGlyAlaValleuGluValHisAsnLysAlaSerLeu 118
QY 417 CACCACCTTTTACTACCAT---AACCAAGAAGTGAAGCAATTGTGGAATTTT---GCTGAT 470
DB 118 hrProLeuLeuLeu-SerileThrLysArgSerGluGlnleuValGluPheLeuLeuile 137
QY 471 AAAATGCAATCGAATCGGTTA---ATAAGTTAAATGCCACACCTCATGCTGCTGAT 527
DB 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVal 157
QY 528 G---TCTGGATCATCAGATAGTTGTCATGCTTCTTAGCAAAA---TGTGACGCTTT-- 579
DB 157 lCysHisGlySerSerGluLeuValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
QY 580 -GTGAGATATAGTGAATCACTGACAGACAT---ATGCTGTACTGTGGATTTCTCACA 635
DB 177 alaAspIle-CysGlyValThrAlaGluHisTrpAlaValThrCysGlyPheHisHis 197
QY 636 TTCATACAAA---TTATGAATATATACGAAATATCAAAATATCA---TCAAAATACCAA 689
DB 197 leHisGluGlnIleMet-GlutryIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAGGAACCTGAGGAACCTGATGAGCTGCACCT---GGCGAAAGACAC 743
DB 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
QY 744 CTGACACGCTGAAAGCT---GGTGGAAAACACCTGATGAGCTGCACCT---GGTGA 797
DB 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
QY 798 AAGACACCTGACAGGCTGAAAGCT---GGTGGAAAACACCTGATGAGCTGCATCT--- 851

```

```

Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
QY 852 TGGTGGAGGAACATCTGACAAATTCATGTTGGAGAAAGG---ACATCTGGAAGTTTCGA 908
Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
QY 909 AGCTCAGC---AGAAAACACCTAGGAATACGATCC---TGCAAAAACACATCTCAGA 962
Db 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
QY 963 AATTACG---GCCAGCAAGAGACCTAGAGATCGCAGGGAGAA---AAAGAAGAC 1016
Db 315 yspheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
QY 1017 ACACCTAGGGAATATGAGTCCGAAAAGAAACACTGAGAAA-----TTA 1061
Db 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
QY 1062 CTTGGGCGAGAAAAGAGACCTTAGGAAGCCGATGGGAGAAAAGAAA---CCTGTGAAA 1118
Db 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
QY 1119 GATGATGCT--GCAAGAGTACATCTAATAAACTAAAG---TTTGAAAAGGAAATCT 1174
Db 372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
QY 1175 AAGATGTT--GCATGTCCCAAAAG---AATCATCTCAAAAGCAAGTCCCATGATCAGAG 1230
Db 392 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg 411
QY 1231 ---TCCCATCATCCAAACAAGGGAAGATCAAGAATATCTTT--GATTTCTCGGCTCT 1286
Db 412 PheProSerGluSerLysGlnGluAspGluGluLysThrSerCysAspSerArg---Ser 430
QY 1287 TTG-----AGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAGT 1337
Db 431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
QY 1338 AATGAGATAAATAAGAAAG---TAGAAAAGCTCTTAAAGCCATCTGCTTCAAGCCGCCA 1394
Db 451 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 470
QY 1395 TTGAAGACAAA---CTCTTTCCAAATAAGCCTTTG-----AATGAAGATGAACA 1442
Db 471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
QY 1443 AACATTGAAGCAGATCCGTG--TTCCACCG---AATCCAAACAAGGACTATAAGAAAAT 1498
Db 489 ThrLeuArgAlaAspProMetPheProProGluSerLysGlnLysAspTyrGluGluAsn 508
QY 1499 TCTTGGGATTCAGAGTCTCTG---GAGACTGTTTACAGAAGG---ATTGTGTTTACCAA 1552
Db 509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy 528
QY 1553 GCCTACATCAAAAAGAAATAGATAAAA---AATGGAATA---TAGAAGAGTCCCTAAT 1606
Db 528 salThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 548
QY 1607 AAATGCTGCTTCTGAGGCTACCTCGGAAT---GAAATTTCTATTCACCTAAAGCCT--- 1660
Db 548 yAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
QY 1661 AGAATTGAAGACATCAAACTTTCAAGCGAGCCTCC---GGGAAGCCTCTGCT---TC 1714
Db 568 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl 587
QY 1715 GAGCTCCACTGAATCAAAAGTCTGTCCTCCAAATAAGCCTTGG-----AATGAAA 1765
Db 587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
QY 1766 ATGAAAAACATGGAGCGAGATGAGAA---CTCCCATCAGATCCAAACAAAGACTATAA 1822

```

Db 607 ---GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyrG 625
Qy 1823 GAAAA---TTCTGGATCTGAGTCTCTG---TAGACTGTTTTCACAGAAGGATTG-TGT 1875
Db 625 LuGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
Qy 1876 TTACCAAGG---CTGCGCTCAAAAGAAATAGATAAATAATGGAATAATAGAGGGT-- 1929
Db 645 LeuProLysAlaAlaHisGlnLysGluLeuLeuAspLysLeuGluGlySer 664
Qy 1930 CCCTGTTAAAGTGTCTTCTGAGCTAACT---GGGAATGAAAGTCTTATTCACACTAAA 1986
Db 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
Qy 1987 GCCTAGA---ATTGATGACATGCAAACTTCAAAG---CAGCGCTCCGACAGCACTGTC 2040
Db 685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
Qy 2041 CTTCGAGCTCCATTGAAA---TCAAAAGTGTGTCACAAATAAAC---CTTGGAAATTAAGA 2094
Db 704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
Qy 2095 ATGAACAAACATTGAGGAGATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTA 2148
Db 724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
Qy 2149 TGAGAAAGTCTTGGGATCTGGAGTCTCTG---TAGACTGTTTCCAGAGGATGGTG-- 2203
Db 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
Qy 2204 -TTTACCAGCTACATCAAAAGAAATAGATAAAA---TAATGGAAATAGAGAGTC 2259
Db 763 sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe 783
Qy 2260 TCT-GATAATCAGGTTTTCAGGCTCCCT---GAGAATGAAAGTCTTATTCACATAAA 2315
Db 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy 803
Qy 2316 GCCT---TGAATTGATGGCATGCAAACTTTCAAAGCAGGCTCCCG---AGAGCCATCTG 2369
Db 803 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 823
Qy 2370 CTTCGAGCTGCATTGAAAT---GAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
Db 823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
Qy 2424 AATCAACAACATTGAGACATCAGAT---TTCCCTTCAGAAATCAAAACAAAGAGG 2477
Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 861
Qy 2478 TTGAGAAA---TTCTGGATCTGAAGTCT---CCGTGACTGTTTCACAGAAGGATGG 2531
Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
Qy 2532 TGTGTACC---CAGCTACACAAAAGAAAGATGATAAAT---AGTGGAAATTAAGAAG 2585
Db 881 alCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
Qy 2586 ATTCACTAGCTCATCAAAAT---CTTGTGACAGTTTCAT---CTTGTGAAAGCGCAAGGAA 2639
Db 901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL 921
Qy 2640 CTTCAAAAGACATCTGGAACA-CGTACAGAAATGGAACAAT---GAAAAGAGATTIT 2695
Db 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 941
Qy 2696 GTGTACTGAAAAAGAAC---TGTCAGAACAAAGAAATAAATACACAGTAGAGAA---CCA 2749
Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 960
Qy 2750 AAAGTTAAATGGGAACAGAG---TCTGCAGTGTAGATTGACTTTAAACCAA---GAAAAG 2803
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980

Qy 2804 AGAAGAGAAATGCCGATATATTAAAGAAAAA-----TTAGGAAGAAATTAG 2851
Db 981 LysArg-ArgAsnAlaAspIleLeuAsnGlnLysIleArgGluGluLeuGlyArgIle-G 1000
Qy 2852 AAGAATCGAGCAGCAGATAGAAAGAGTGTAGAAGTGAACACACAACTTTGAAG---GCTCTC 2908
Db 1000 LuGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
Qy 2909 AGATCAAG---ATAGAATGAAGGTGTAGAAAGTAATTTCAATAGGTTTCTCA---AC 2962
Db 1018 ArgGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
Qy 2963 TCATGAATGAAA---ATTATCTTACATGAATTCATGTTGTAAGAAAGGATTCGCA-- 3017
Db 1037 rHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaLe 1057
Qy 3018 -TGCAAACTGAATAGCCACACTA---AACACCAATCCAGGAAAGGAAAAATAATC-T 3072
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
Qy 3073 TTGAGGACTTAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
Qy 3129 AAAGGAATCATTTACTAAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA 3182
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
Qy 3183 GCTGAGAAACAATGCTCATTTCTAAATTTGAGGAAACACAGACAAAGAAA-----CTAGA 3236
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGI 1136
Qy 3237 GCAGAAATTGAATA-CACCATCCTGACTGGCTTCGCTGTACAAGC-----CA 3283
Db 1136 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
Qy 3284 TGATCAATTTGACATCAAGAAAAAAGTAAGA---ACCTGCTTCCACATTGAG---GAGAT 3337
Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
Qy 3338 GCTGTTTCAAGAAATGAATGTTTGATGTG---AGAGTACGATATTACATGATGCT 3394
Db 1173 aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAsnGluValL 1193
Qy 3395 CCATCACCACCTTTC---TGAGCTCAAGAAATCCAAACCTCAAAATTTATCT---CAAT 3448
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
Qy 3449 TAGCAGGAGATGCTTAAGAGA---AATCATTTGGTTTCAGAACATCACAAAGAGA---CA 3502
Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGI 1232
Qy 3503 ACGTGAACACAGTGTCAAT---GAAGGAAGCGAACACATGTATCAAAAC-----GACAA 3553
Db 1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
Qy 3554 GATAATGGAAACAAACACCTGAACAGCAGAGTCTCTAGTCAGAAATATTTCACATCAAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
Qy 3614 CAAAATTTGCTTTCACAGCAATT---AGTCATGCACATAGAAAGCTGACACAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
Qy 3670 CAGATAACAAT-----GATATTTCATTTCTTTG---AGAGGAAATGCAACAC 3711
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
Qy 3712 ATCTCTTAAACAGAAAAAGAGGAGATATT---AATTACAATACCATTTTAAAAACCGTAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327

```

QY 3769 ATT---CAATATGAAAAAGACAAA 3789
Db 1328 lIeTyrGlnTyrGluLysGluLys 1335

RESULT 2
US-10-007-805-565
; Sequence 565, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuguo
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-565

Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 13 Gaps: 115

US-09-602-362e-22 (1-4115) x US-10-007-805-565 (1-1341)
QY 96 CAACGGAGAGACATCAACCTTAATATACAGAG---CCACAGAGAGCTGCTCTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTCGTCAATGGCCT-----GAGGAAGTAGTAACATTTTC---GGTA 194
Db 22 -----HistTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuVala 38
QY 195 GACAGAAGTGCAGTACGCTTCCTTG---TGGCGAACCAGACACCTCTGTATGAGGC-T 250
Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAla 58
QY 251 TACAAATGACAGGAGGCTTTGC---AAATATTTGATAGATTC-GGTGCCGATATAATC 306
Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTRATGCAACATGGCTTCCATTATGGTTTATAGTGGATTTCAGT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 ---GGTGGCAAACTGGTGC-CATGTCAGATCGAATGTC---CAACAGGCTGCCTCA 416
Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
QY 417 CACCACTTTTACTACCAT---AACGAAAGAGTGAGCAATTTGTGGAATTTT---GCTGAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAATGCAAAATGCGAATGCGTTA---ATAAGTTAAATGCACACCTCATGCTGCTGAT 527
Db 489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508

138 LysAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVal 157
528 G---TCTGGATCATCAGAGTAGTTGTCATGCTTTAGCAAAA---TGTGACGCTTTT-- 579
157 lCysHisGlySerSerGluIleValGlyWeLeuLeuGlnGlnAsnValAspValPheAl 177
580 -GTGCAGATATAGTGAGTAACCTGCAGAACAT---ATGCTGTTACTGTGGATTTCTCACA 635
177 aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIle 197
636 TTTCAATACAAA---TTATGAATATATAGAAAATTTATCAAAATCA---TCAAAATACAA 689
197 lHisGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
690 T---CAGAAGGAACCTCTGCAGGAAACCTGAGGCTGCACCCCT---GGCGGAAGACAC 743
216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGCTGCACCCCT---GGTGA 797
236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
798 AAGACACCTGACAGGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGCTGCATCT--- 851
256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
852 TGGTGGAGAACATCTGACAAATTCATGTTGGAGAAAG---ACATCTGGAAGTTGGA 908
276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 295
909 ACGTCAGC---AGAAAAACACCTAGGAAATACGATCC---TGCAAAAAACACCTCTGAGA 962
295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
963 AATTTAAG---GGCCAGCAAGGAAGACCTAGAGATCGCAGGGAGAA---AAGAAGAC 1016
315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysGluAsp 334
1017 ACACCTAGGGAATATGAGTCCCGAAAAAGAACACTGAGAAA-----TTA 1061
335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
1062 CGTGGCGGAAAAAGGAAGACCTAGGAAGACGATCGGGAAGAAAAAGAAA---CCCTGTAAA 1118
355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
1119 GATGATGCGT-GCAAGAGTACATCTAATAAACTAAAG---TTTGAAGAAAGAAATCT 1174
372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
1175 AAGATGTT-GCATGTGCCAAAG---AATCATCTCAAAAGCAAGTGCATGATCAGAGG 1230
392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
1231 ----TCCATCAGATCCAAACAGGGAAGATGAGATATTTCTTT-GATTCGGGTCTCT 1286
412 PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser 430
1287 TTG-----AAGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAGT 1337
431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
1338 AATGAGATAAATAAGAG---TAGAAAGCTCCTCTAAAGCCATCTGCTCTCAAGCGCCA 1394
451 MetGluIleAsnArgGluValGluGluProLysLysProSerAlaPheLysProAla 470
1395 TTGAAGACAAA---CTCTTTCCAAATAGCCCTTTG-----AATGAAGAAATGAACA 1442
471 lIleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
1443 AACATTAAGAGATCGGTG-TTCCCAACG---AATCCAAACAAAGACTATAGAAAT 1498
489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508

```


Qy	1499	TC TTGGGATCTTAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAA	1552
Db	509	SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys	528
Qy	1553	GGCTACACATCAAAAGAANTAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT	1606
Db	528	SalAThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL	548
Qy	1607	AAAGTGTCCTCTGAAGCTACCTCGGAAT---GAAATTTCTATTCCACTAAAGCCT---	1660
Db	548	ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu	567
Qy	1661	AGAATTGAAGACATGCAAACTTTCAAAAGCAGCCCTCC---GGGAAGCCACTCTGCT---	1714
Db	568	GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG	587
Qy	1715	GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATAAGCCTTGG-----AATGAAAA	1765
Db	587	uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu--	606
Qy	1766	ATGAARAACATGGAGGCAGATGAAA---CTCCCATCAGATCCAAACAAGACNATAAA	1822
Db	607	----GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTrpG	625
Qy	1823	GA AAA---TTCTGGGATCTGGAGTCTCTG---TAGACTGTTTCACAGAAGGATTG-TGT	1875
Db	625	LuGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys	644
Qy	1876	TTACCAAG---CTGCCCTCAAAAGAATAGATAAAATATGGAAATTTAGAAGGTT---	1929
Db	645	LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer	664
Qy	1930	CCCTGTTTAAAGTGGTCTCTGAGCTTAAC---GGGAATCAAAAGTCTTATTCCAACATAA	1986
Db	665	ProValLysAspGlyLeuLeuLysAlaAsnGlyMetLysValSerIleProThrLys	684
Qy	1987	GCCTAGA---ATTGATGACATGCAAACTCAAAAG---CAGGCCCTCCGAGAGCATCTGC	2040
Db	685	Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl	704
Qy	2041	CTTCGAGCCTCCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGGAAATTAAGA	2094
Db	704	apHeGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA	724
Qy	2095	ATGAACAAACATTGAGCAGATGAGAA---CTCCCATCA---AATCCAAACAAGGACTA	2148
Db	724	sngLuuGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr	743
Qy	2149	TGAAAGATCTTTGGGATCTCGAGTCTCTG---TAGACTGTTTCCAGAAGGATGGTG---	2203
Db	744	GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy	763
Qy	2204	-TTTACCAGCGTACACATCAAAAGAANTAGATAAAA---TAATGGAAAATAGAAGTCT	2259
Db	763	sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe	783
Qy	2260	TCT-GATAATGAGTCTTTCTGAAGCTCCCT---GAGAATGAAAGTTCTATTCCACTAAA	2315
Db	783	rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy	803
Qy	2316	GCCT---TGAATTGATGGCATGCAAACTTTCAAGCAGGCCCTCCG---AGAGCCATCTG	2369
Db	803	SalALeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl	823
Qy	2370	CTTCGAGCCTGCATTTGAAAT---GAAAAGTCTGTTCAAAATAAAC---CTTGGAAATTAAG	2423
Db	823	apHeGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys	842
Qy	2424	AATGAACAACTTGAGACACATCAGAT-----TTCCCTTCAGAACTCAAAACAAGAAGG	2477
Db	843	AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys	861

QY	2478	TTGAGAAA---TTCTGGGATTCTGAAGTCT---CCGTGCACTGTGTTTCACAGAAAGGATGG	2531
Db	862	ValGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V	881
QY	2532	TGTTGTACC---CAGGCTACACACAAAAAGAAAGCATAAAT---AGTGGAAAAATTACAAG	2585
Db	881	alCysValProLysalaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA	901
QY	2586	ATTCTAGTACGCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAAGGCAAGGAA	2639
Db	901	spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL	921
QY	2640	CTTCAAAAAGACACTGTGAACA-CGTACAGCAAAATGGAACAAT---GAAAAAGAGTTTT	2695
Db	921	euGlnLysAspHisCysGluLgnArgThrGlyLysMetGluGlnMetCysLysLysPheC	941
QY	2696	GTGTACTCAAAAAGAAC---TGTCAGAACAAAAAGAAATAAATCACACTAGAGAA---CCA	2749
Db	941	ysValLeuLysLysLeuSerGluAlaLysGluLysSerGln-LeuGluAsnGln	960
QY	2750	AAAGCTTAATGGGAACAAGAG---TCTGCAGTGTAGATTCACTTAACCAA---GAAAG	2803
Db	961	LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu	980
QY	2804	AGAGACAGAGAAATGCCGATATATTAAAGAAAA---TTAGCAAGAAATTAG	2851
Db	981	LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluLysGlyArgIle-G	1000
QY	2852	AAGAATCGAGCAGCAGTAGAAGAGTTAGAGTGAACAACAACACTTGAAAG---GCTCTC	2908
Db	1000	luGluGlnHisArg-----LysgluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu	1017
QY	2909	AGATACAAG---ATAAGAAATTGAAGGTGTAGAAGTAATTTGAATAGTTTCTCA---AC	2962
Db	1018	ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisThr	1037
QY	2963	TCATGAATGAAA---ATTATTCTTACATGAAATTCGATCTTGAAAAAGAAATTCGCA--	3017
Db	1037	rHisLysAsnGluAsnTrpLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe	1057
QY	3018	-TGCAAACTCGAATAGCCACACTA---AACACCAATCCAGGAAAGGAATAAATC-T	3072
Db	1057	tLeuLysLeuGluIlealaThrLeuLysHis-GlnTrpGlnGluLysGluAsnLysTrp	1077
QY	3073	TTGAGGACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATCACCTAAAACTGA	3128
Db	1077	heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL	1097
QY	3129	AAAGGAATCATTACTAAAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA	3182
Db	1097	ysGluGluSerLeuThrLysArgAlaSerGlnTrpSerGly-GlnLeuLysValLeuIle	1116
QY	3183	GCTGAGAAAAACAATGCTCATTTCTTAATTGAGAAAAACACAAAGAAAA---CTAGA	3236
Db	1117	AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleuGlu	1136
QY	3237	GGCAGAAATTGATA-CACCATCTGACTGGCTTCGCTGTACAAGC-----CA	3283
Db	1136	uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi	1153
QY	3284	TGATCAATTGTGACATCAACAAAAAGTAAGA---ACCTGCTCCACATTCAG---CAGAT	3337
Db	1153	sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl	1173
QY	3338	GCTGTTTTCAGAAAAATGAAATCTTGATGTG---AGAGTACGATATTAAACAATGAGTCT	3394
Db	1173	aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTrp-AsnAsnGluValu	1193
QY	3395	CCATCACCACCTTC---TGAGCTCAAGGAATTCAAAACCTTAAATATCT---CAAT	3448
Db	1193	euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnTr	1213
QY	3449	TAGCAGGAGATGCTCAAGAGA---AATAACATTTGTTTCAGAACATCAAAAGAGA---CA	3502

Db 1213 yralagiYasp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGI 1232
QY 3503 ACGTGAAGAACAGTGTCAAT---GAAGGAAGCGAACACATGTATCAAAAC-----GACAA 3553
Db 1232 nargGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY 3554 GATAATGGAACAAACACTGAACAGCAGAGTCTTCTAGTCAGAAATTTTCACTACAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAAATGTGCTTCACAGCAAT---AGTCATGCATAGAAAGCTGACAAACAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATAACAAT-----GATATTCATTTCTTG---AGAGGAATATGCAACAC 3711
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCCTTAAAGAGAAAGAGGAGATATT---AATTACATATCATTATTAACACCGTAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY 3769 ATT---CAATATGAAAAAGAAAA 3789
Db 1328 ileTyrGlnTyrGluLysGluLys 1335
RESULT 3
US-10-076-622-565
; Sequence 565, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-076-622-565
Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 14 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-076-622-565 (1-1341)
QY 96 CAAAGGAAGAGACATCAACCTTAATATACAAGAG---CCGAGAAGAGACTGTCTTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGTGCTCAATGGCCT-----GAGGAAGTAGTAACATTTC---GGTA 194
Db 22 -----HisTrpAlaCysValAsnGlyHisGlnGluValThr-PheLeuVala 38
QY 195 GACAGAAGTGCAGTGTGCTCTTG---TGCGCAACACGACACCTCTGTATGAAGGC-T 250
Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACAATGACAGGAGGCTTTGC-----AAATATTGATAGATTC-GGTGCCGATATAATC 306
Db 58 euGlnCysHisGlnGlnAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78

QY 307 TCGTG---ATGTGTATGCAACATCGCTTCCATTATGCGTTTATAGTGGATTTTGTCACT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 --GGTGGCAAACTGTGTCTC-CATGCTGCAGTATCGAAGTGC---CAACAAGCTGCTCA 416
Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118
QY 417 CACCACTTTTACTACCAT---AACGAAGAAGTGCAGCAATTTGTGAATTTT---GCTCAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle 137
QY 471 AAAAATGCAATCGAATGCGTTA---ATAAGTTAAATGCACACCCTCATGCTGCTGTAT 527
Db 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
QY 528 G---TCTGATCATCAGAGATAGTTGCAATGCTTCTTACCAAAA---TGTGACGCTTTT-- 579
Db 157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 177
QY 580 -CTGCAGATATAGTGAAGTAACTGCAGAACAT---ATGCTGTACTGTGGATTCTTCTACA 635
Db 177 alaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIle 197
QY 636 TTCATAACAAA---TTATGATATATACGAAATATCAAAAATCA---TCAATATACCAA 689
Db 197 leHisGluGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAAGGAACCTCTGCAGAAACCTGTAGAGCTGCACCT---GGCGGAAGACAC 743
Db 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 236
QY 744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGTAGAGCTGCACCT---GGTGGGA 797
Db 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
QY 798 AGACACCTGACAGGCTGAAAGCT---GGTGGAAAAACACCTGTAGAGCTGCATCT--- 851
Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
QY 852 TGTGTGAGGAACATCTGCACAAATTCATGTTGGAGAAAG---ACATCTGGAAGTTTCCA 908
Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheG 295
QY 909 ACCTCAGC---AGAAAAACACCTAGAAATACAGATCC---TGCAAAAAACACATCTGAGA 962
Db 295 uGln-SerAlaGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
QY 963 AATTTACG---GGCCAGCAAGGAAGACCTAGAAGATCGAGGAGAA---AAAGAAGAC 1016
Db 315 yspPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
QY 1017 ACACCTTAGGGAATATGACTCCGAAAAAGAAACACATGAGAAA-----TTA 1061
Db 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
QY 1062 CGTGGCAGAAAAAGGAGACCTAGGAAGACGATCGAGAAAAAGAAA---CCCTGTAAA 1118
Db 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
QY 1119 GATGATGCGCT-GCAAGAGTAACATCTATAAACTAAAG---TTTGTAAAAAGGAATCT 1174
Db 372 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
QY 1175 AGATGTTT-GCATGTCCAAAAAG---ATCATCTACAAGCAAGTGCATCATCAGAGG 1230
Db 392 LysMetIleAlaCysProThrLysGluSerThrLysAlaSerAlaAsnAspGlnArg 411
QY 1231 ---TCCCATCATCCCAAGGGAAGATGAAGATATTTCTTT-GATTCTCGGTCTCT 1286
Db 412 PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser 430

1136 ualaGluileGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
3284 TGATCAATTGGACATCAAGAAAAGTAAGA---ACCTGCTTCCACATTGAG---GAGAT 3337
1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
3338 GCTGTTTGCAGAAAGAAATGAATGTTGATGCG---AGAGTACGATATTAAACAATGAGTCT 3394
1173 aCysLeuGlnArgLysMetAsnValAspValSerThrIleTyr-AsnAsnGluValL 1193
3395 CCATCACCACCTTC---TGAGCTCAAGGAATCAAAACCTAAATAATTTATCT---CAAT 3448
1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
3449 TAGCAGGAGATGTTCTAAGAGA---AATACATGTTTTCAGAACATCAACAAGAGA---CA 3502
1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
3503 ACCTGARAACAGTGTCAAT---GAAGGAAGCGAACACATGATCAAAAC-----GACAA 3553
1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
3554 GATATGGAACAAACACCTGACAGCAGAGTCTTAGTCAGAAATTTTCAACTACAG 3613
1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlnS 1272
3614 CAAAATTTGGCTTCACAGCAAT---AGTCATGCACATAGAAAGCTGACACAAAG 3669
1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
3670 CAGATAACAAT-----GATATTCTATTTCTTG---AGAGGAATGCAACAC 3711
1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
3712 ATCTCCTTAAAGAGAAAAGAGGAGATATT---AATTACATACATTTAAACCGTAT 3768
1308 lsLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
3769 ATT---CAATATGAAAAGAGAAA 3789
1328 lietyrGlnTyrGluLysGluLys 1335
RESULT 4
US-10-177-293-334
; Sequence 334, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; CURRENT APPLICATION NUMBER: US/10/177,293
; FILE REFERENCE: MRI-038
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572

PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 334
LENGTH: 1341
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-293-334
Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 14 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-177-293-334 (1-1341)
QY 96 CAAAGGGAAGACAGATCAACCTTAATATACAAGAG---CCGAGAAGAGACTGTCTTAAC 152
DB 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGTCTCAATGCGCT-----GAGGAAGTAGTAACATTC---CGTA 194
DB 22 -----HisTrpAlaCysValAsnGlyHisGluValValThr-PheLeuVala 38
QY 195 GACAGAGTGCAGCTGACGCTCTTG---TGGCGAACACGACACCTCTGTATGAGGC-T 250
DB 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACAATGCCACAGAGGCTTTGC---AAATATTGATGATTC-GTGCCGATATAATC 306
DB 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGCTTCCATTATGCGTTTATAGTGGATTTTTCAGT- 362
DB 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 --GGTGCACAACTGTGTC-CATGCTGCAGTATCGAAGTGC---CAACAAGCTCGCTCA 416
DB 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeu 118
QY 417 CACCATTCTTACTACCAT---AACGAAAGAGTGCATTCGTGAATTTT---GCTGAT 470
DB 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeu 137
QY 471 AAAAATGCAAAATCGCAATCGTTA---ATAAGTTAAATGCACACCCCTCATGTGCTGAT 527
DB 138 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 157
QY 528 G---TCTGGATCATCAGATAGTTCATGCTTCTTAGCAAAA---TGTGAGCTGCTTT-- 579
DB 157 lCysHisGlySerSerGluIleValGlyMetLeuLeuGlnAsnValAspValPheAl 177
QY 580 -GTGCAGATATAGTGCAGTAACCTGCAAGAACAT---ATGCTGTACTGTGGATTTCTACA 635
DB 177 alaaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIst 197
QY 636 TTCATAACAAA---TTATGAATATATACGAAATATCAAAATCA---TCAAAATACCA 689
DB 197 leHisGluGlnIleMet-GlutryIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 216
QY 690 T---CAGAAGGAACCTCTGCAGAAACCTGATGAGGCTGCACCT---GGCGGAAGACAC 743

Db 216 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrp 236
Qy 744 CTGACAGCGTGAAAGCT---GGTGGAAAAACACCTGATGAGGTGCACCCCT---GGTGGGA 797
Db 236 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 255
Qy 798 AAGACACCTGACAGCGTGAAAGCT---GGTGGAAAAACACCTGATGAGGTGCATCT--- 851
Db 256 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 275
Qy 852 TGGTGGAGGACATCTGACAAATCAATGTTGGAGAAAG---ACATCTGGAAGTTCGA 908
Db 276 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGlu 295
Qy 909 ACCTCAGC---AGAAAAACACCTAGGAAATPACGATCC---TGCAAAAAACATCTGAGA 962
Db 295 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 315
Qy 963 AATTTCAG---GGCCAGCAAGAGGACCTAGAGATCGCAGGGAGAA---AAGAGAGAC 1016
Db 315 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 334
Qy 1017 ACACCTAGGGAATPATGAGTCCCGAAAAAGAAACACTGAGAAA-----TTA 1061
Db 335 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 354
Qy 1062 CGTGGGAGAAAAGGAGACCTAGGAGAGCATGGGAGGAAAAGAAA---CCCTGTAAA 1118
Db 355 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 371
Qy 1119 GATGGATCGCT-GCAAGAGTAACATCTAATAAACTAAAG---TTTTGAAAAAGGAATCT 1174
Db 372 ThrGlyCysValAlaAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 391
Qy 1175 AAGATGTT-GCATGTCACAAAAG---AATCACTCAAAAGCAGTGCCATGATCAGAGG 1230
Db 392 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg 411
Qy 1231 ---TCCCATCAGATCCAAAACAAGGAAGAGTGAAGATATTCTTT-GATTCTCGGGTCTCT 1286
Db 412 PheProSerGluSerLysGlnGluGluAspGluGluTrpSerCysAspSerArg---Ser 430
Qy 1287 TTG-----AAGTTCCTCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAAGT 1337
Db 431 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyrGlnLysVal 450
Qy 1338 AATGAGATTAATAAGAG---TAGAAGCCTCTTAAAGCCATCTGCTTCAAGCCGCCA 1394
Db 451 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 470
Qy 1395 TTGAAAGCAAAA---CTCTTCCAAATAAGCCTTTG-----AATGAAGATGAACA 1442
Db 471 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 488
Qy 1443 AACATTGAAGCAGATCCGTG-TTCCCAACG---AATCCAACAAAGGACTATAGAATAAT 1498
Db 489 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn 508
Qy 1499 TCTGGGATTCTAGACTCTCTG---GAGACTGTTTACAGAAAG---ATTGTGTTTACCAA 1552
Db 509 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys 528
Qy 1553 GGCTACACATCAAAAAGATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT 1606
Db 528 salAThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 548
Qy 1607 AAGTGTGCTCTCAAGCGTACCTCGGAAT---GAAATTCTATTCCACTAAGGCCT--- 1660
Db 548 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 567
Qy 1661 AGAATTGAAGACATGCAAACTTTCAAAAGCGAGCGCTCC---GGGAAGCCATCTCTCT---TC 1714
Db 568 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGlu 597

Qy 1715 GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTTGG-----AATGAAAA 1765
Db 587 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu-- 606
Qy 1766 ATGAAAAACATGGAGCGCAGATGAGAA---CTCCCATCAGATCCAAACAAAGGACTATAA 1822
Db 607 ---GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyrG 625
Qy 1823 GAAAA---TTCTGGGATCTACTGAGTCTCTG---TAGACTGTTTTCACAGAGGATTG-TGT 1875
Db 625 luGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 644
Qy 1876 TTACCAAGG---CTCGGCTCAAAAAGATAGATATAAATAATGGAAAAATTAGAGGGT--- 1929
Db 645 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLys 664
Qy 1930 CCTGTAAAGTGGTCTTCTGAGGCTAACT---GGGAATGAAGTTCATTATCCAACTAAA 1986
Db 665 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 684
Qy 1987 GCCTAGA---ATTGATGACATGCAAACTTCAAG---CAGGCTCCCGAGAGCCATCTGC 2040
Db 685 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 704
Qy 2041 CTTTCGAGCTCCATTGAAA---TCAAAAGTCTGTCCCAATAAAC---CTTGAATTAAGA 2094
Db 704 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 724
Qy 2095 ATGACAAAACATTGAGCAGATGAGAA---CTCCCATCA---AATCCAAACAAAGACCTA 2148
Db 724 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr 743
Qy 2149 TCAGAAAGTCTTTCGGATTCGAGTCTCTG---TAGACTGTTTTCACAGAGTGGTG--- 2203
Db 744 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 763
Qy 2204 -TTTCCACGAGGTACACATCAAAAGAAATAGATAAAA---TAATGAAAAATAAGAGAGTC 2259
Db 763 slenProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluLys 783
Qy 2260 TCT-GATTAATGAGGTTTCTGAGAGCTCCCT---GAGAAATGAAGTCTTATTCCTCAATAA 2315
Db 783 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrly 803
Qy 2316 GCCT---TGAATTGATGCGATGCAAACTTCAAGCAGCGCTCCCG---AGAGCCATCTG 2369
Db 803 salalLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 823
Qy 2370 CTTGAGCGCTGCATTGAAAT---GAAAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAG 2423
Db 823 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 842
Qy 2424 AATGAACAACATTGACAGATCAGAT-----TTCCCTTCAGANTCAAAACAAAGAGG 2477
Db 843 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLys 861
Qy 2478 TTGAGAAAA---TTCTGGGATCTCAAGTCT---CCGTGGACTGTTTTCACAGAGGATCG 2531
Db 862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
Qy 2532 TGTGTACC---CAGGCTTCACAAAAAAGAGGATAAAAT---AGTGGAAATTAGAG 2585
Db 881 aiCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
Qy 2586 ATTCTACTAGCTTATCAAAAT---CTTGTACAGTTTCAT---CTTGTGAAGCAAGGAA 2639
Db 901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGluL 921
Qy 2640 CTTCAAAAAGACATCTGAACA-CGTACAGGAAATGGAACAAT---GAAAAAGAGTTT 2695
Db 921 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941

```
QY 2696 GTGTACTGAAAGAAC---TGTGACACAAAGAGAAATAATCACAGTAGAGAA---CCA 2749
Db 941 ysValLeuLysLysLysLeuSerGluAlaLysGluLeLysSerGln-LeuGluAsnGln 960
QY 2750 AAAGTTAAATGGACACAGAG---TCGTGACGTGTAGATTGACCTTAAACCAA---GAAAG 2803
Db 961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
QY 2804 ACAAGAGAGAAATGCCGATATATTAAAGAAAAA---TTAGGAGAAATTAG 2851
Db 981 LysArg-ArgAsnAlaAspLeuLeuAsnGluLysLeuLeuArgGluLeuGlyArgIle-G 1000
QY 2852 AAGATCGAGAGCAGCATAGAAAGATTAGAGTTGAGAGTGAACACAACTTGAAG---GCTCTC 2908
Db 1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY 2909 AGATACAAAG---ATAAGAAATTGAAGGTGTAGAAAGTAAATTTGAATAGTTCTCA---AC 2962
Db 1018 ArgIleGlnAspLeuGluLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
QY 2963 TCATGAAATGAAA---ATTATCTTACATGAAATTCATGTTGAAAAGGAATGCCA--- 3017
Db 1037 rHisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY 3018 -TGCAAACTGGAATAGCACACTA---AACACCAATCCAGAAAAGAAATAAATC-T 3072
Db 1057 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1077
QY 3073 TTGAGGACTTAAGATTTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAATCTGA 3128
Db 1077 heGluAspIleLysIleLeuLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1097
QY 3129 AAAGAAATCACTATAAG---GGATCTCAATATGTTGGCAGCT---AAAGTCTGATA 3182
Db 1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1116
QY 3183 GGTGAGAAACATGCTCTATTCTAAATTCAGGAAAACACAGAAAGAAAA---CTAGA 3236
Db 1117 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1136
QY 3237 GGCAGAAATGAATA-CACATCTCTGACTGCTTCGTGTCACAAGC-----CA 3283
Db 1136 uAlaGluIleGluSerHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY 3284 TGATCAATTTGTGACATCAAGAAAAAGTAAGA---ACCTGCTTCACATTCAGAG---GAGAT 3337
Db 1153 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY 3338 GCTGTTTCCAAAGAAATGAATGTTGATGTG---AGAGTAGATATTAAACAATGAGTGCT 3394
Db 1173 acysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1193
QY 3395 CCATCACCACITTC---TGACGTCAAAGGAAATCCAAACCTAAATTAATCT---CAAT 3448
Db 1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1213
QY 3449 TAGCAGGAGAGTCTCTAAGAGA---AATACATTGGTTTCAGAACATCAACAAGAGA---CA 3502
Db 1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
QY 3503 ACCTGAAAACAGTGTCAAT---GAAGGAGGAGACACATGATCAAAAC-----GACAA 3553
Db 1232 rArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
QY 3554 GATAATGGAACAAACACCTCAACGACGAGTCTCTAGTCAGAAATTTTCAACTACAAG 3613
Db 1252 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
QY 3614 CAAAATTTGGCTTCACAGCAAT---AGTCATGCACATAGAAAGCTGACACAAAG 3669
Db 1272 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1288
QY 3670 CAGATAACAAT-----GATATTCTTTCTTCTG---AGAGGAAATGCAACAC 3711
```

```
Db 1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
QY 3712 ATCTCTCTAAAGAGAAAAAGAGAGATATT---AATTACAATACCATTTAAAAACCGPAT 3768
Db 1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
QY 3769 ATT---CAATATGAAAAAGAGAAA 3789
Db 1328 IleTyrGlnTyrGluLysGluLys 1335
RESULT 5
US-10-124-805-565
; Sequence 565, Application US/10124805
; Publication No. US20030166022A1
; GENERAL INFORMATION:
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 21021.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 565
; LENGTH: 1341
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-124-805-565
Alignment Scores:
Pred. No.: 8,89e-110 Length: 1341
Score: 1234.00 Matches: 679
Percent Similarity: 59.18% Conservative: 146
Best Local Similarity: 48.71% Mismatches: 376
Query Match: 17.41% Indels: 207
DB: 14 Gaps: 115
US-09-602-362E-22 (1-4115) x US-10-124-805-565 (1-1341)
QY 96 CAAAGGAAAGACAGACATCACTTAATATACAGAG---CCAGAGAGACTGTCTTAAC 152
Db 3 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 21
QY 153 TGGGCTGTGTCATTCGCTCCT-----CAGGAAGTAGTAACATTTC---GGTA 194
Db 22 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 38
QY 195 GACAGAGAGTCCAGCTGAGTCTCTTG---TGCGAACACGGACCTCTGTATGAGGC-T 250
Db 38 spArgLysCysGlnLeuAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL 58
QY 251 TACAATGCCACAGGAGGCTTTCG---AAATATTGATAGATTTC-GGTCCGATATAAATC 306
Db 58 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 78
QY 307 TCGTG---ATGTGTATGCAACATGCTTCATTATTCGTTTATAGTGATTTTGTGAGT- 362
Db 78 euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluIleLeuSerV 98
QY 363 ---GGTGGCAAACTGTGTC-CATGGTGAGTATCCAGTGC---CAACAAGGCTCCCTCA 416
Db 98 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 118
QY 417 CACCACCTTTTACTACCAT---AACGAAAGAAAGTAGCAATTGTGGAATTTT---CCTGAT 470
Db 118 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle 137
QY 471 AAAAATGCAAAATGCAATCGGTTA---ATAAGTTAAATGACACACCTCATGCTGTGTAT 527
Db 138 LysAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaValA 157
```

QY	528	G	---	TC	TG	AT	CAT	CAG	AG	TAG	TT	GC	AT	GC	TCT	CT	TAG	CA	AAA	---	TGT	G	AC	GC	T	T	T	---	579																															
Db	157	I	C	Y	S	H	I	S	G	L	Y	S	E	R	G	L	U	L	E	V	A	G	I	G	L	U	G	L	N	V	A	L	A	S	P	V	A	L	P	H	E	A	L	177																
QY	580	-	GT	C	A	G	A	T	A	T	A	G	T	G	G	A	G	T	A	C	T	C	C	A	A	A	---	-	A	T	G	T	T	A	C	T	G	T	G	G	A	T	T	C	T	C	A	C	A	635										
Db	177	a	A	A	a	s	p	i	l	e	-	C	y	s	e	G	L	y	v	a	l	t	h	r	a	l	a	l	u	H	i	s	t	y	r	a	l	a	v	a	l	t	h	r	C	y	s	e	G	L	y	p	h	e	i	s	i	197		
QY	636	T	T	C	A	T	A	C	A	A	---	-	-	-	T	T	A	T	A	T	A	T	A	C	A	A	A	A	A	T	T	A	C	A	A	A	T	C	A	---	-	-	T	C	A	A	T	A	C	A	A	689								
Db	197	l	e	H	i	s	G	L	U	G	L	e	-	G	l	u	T	-	I	e	A	r	g	i	y	S	e	r	-	L	y	S	a	s	n	H	i	S	G	L	i	n	A	s	n	T	h	r	A	s	216									
QY	690	T	---	C	A	G	A	G	A	C	T	C	C	A	G	A	A	C	T	G	A	G	G	T	C	C	A	C	---	-	-	-	G	G	C	G	A	A	A	G	A	C	A	C	743															
Db	216	n	P	r	O	G	L	U	G	L	y	T	h	r	S	e	r	A	l	a	G	L	y	T	h	r	P	r	O	a	s	p	G	L	U	a	l	a	-	P	r	O	L	e	U	a	l	a	g	l	u	a	r	g	t	h	r	P	236	
QY	744	C	T	G	A	C	A	C	G	C	T	G	A	A	G	C	T	---	-	-	-	G	G	T	G	A	A	A	A	A	A	C	A	C	T	G	A	T	G	A	G	G	T	G	C	A	C	C	---	-	-	-	G	G	T	G	G	A	797	
Db	236	r	O	a	s	p	T	h	r	A	l	a	G	L	e	U	S	e	r	V	a	l	G	L	U	y	S	t	h	r	P	r	O	a	s	p	G	L	U	a	l	a	-	P	r	O	L	e	U	V	a	l	g	l	u	255				
QY	798	A	A	G	A	C	C	T	C	A	C	G	C	T	C	A	A	G	C	---	-	-	-	G	G	T	G	A	A	A	A	A	C	A	C	T	G	A	T	G	A	G	G	T	G	C	A	T	---	-	-	-	851							
Db	256	A	r	g	t	h	r	P	r	O	a	s	p	T	h	r	A	l	a	G	L	e	U	S	e	r	V	a	l	G	L	U	y	S	t	h	r	P	r	O	a	s	p	G	L	U	a	l	a	-	A	l	a	s	e	r	L	e	U	275
QY	852	T	G	T	G	G	A	G	A	C	A	T	C	T	C	A	C	A	A	T	T	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	908										
Db	276	V	a	l	-	G	L	U	G	L	y	T	h	r	S	e	r	A	s	p	L	y	S	i	l	e	G	L	i	n	C	y	S	L	e	U	G	L	U	y	S	a	l	a	t	h	r	S	e	r	G	L	y	S	p	H	e	G	L	295
QY	909	A	C	G	T	C	A	C	---	-	-	-	A	G	A	A	A	A	C	A	C																																							

Qy	1499	TCTTGGAGTCTAGAGTCTCTG---	GAGACTGTTTACAGAGG---	ATTGCTGTTTACCAA	1555			
Db	509	SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLys	528	GGCTACACATCAAAAGAAATAGATAAAAA--	AATGGAAAA--TAGAAGAGTCCCTTAAT	1606		
Qy	1553	GGCTACACATCAAAAGAAATAGATAAAAA--	AATGGAAAA--	TAGAAGAGTCCCTTAAT	1606			
Db	528	sAlaThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL	548	ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu	567			
Qy	1661	AGAATTCAAGACATGCAAACTTTCAAGCGAGCTCC-	GGGAAGCCATCTGCT--	TC	1714			
Db	568	GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG	587	1715	GAGCTCCACTGAAATCAAAAGTCTGTCCTCAAAATAAGCCCTTGG-----	AATGAAAA	1765	
Qy	1715	GAGCTCCACTGAAATCAAAAGTCTGTCCTCAAAATAAGCCCTTGG-----	AATGAAAA	1765	587	uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu--	606	
Db	1766	ATGAAAACATGAGCAGATGAGAA--	CTCCCATCAGAAATCAAAACAAAGACTATAA	1822	607	---GlnThrTrpArgAlaAsp-GluLeuLeuProSerGluSerLysGlnLysAspTyrG	625	
Qy	1823	GA AAA--TTCTGGGATCTCGAGTCTCTG---	TAGACTCTTTTACAGAGGATTTG-TGT	1875	625	luGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys	644	
Db	1876	TTACCAAG--	CTGCGCTCAAAAAGAAATAGATAAAATAATGGAATATTAGAAGGT---	1929	645	LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer	664	
Qy	1930	CCCTGTTAAAGTGTCTTCTGAGCTAACT---	GGGAATGAAAGTCTTATTCACAACTAAA	1986	665	ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys	684	
Db	1987	GCTTAGA--	ATTGATGACATGCAAACTTCAAAG---	CAGCGCTCCCGAGAGCCATCTGC	2040	685	Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl	704
Qy	2041	CTTCGAGCTCCCATTTGAAA--	TCAAAAGTCTGTCCAAAATAAAAC---	CTTGGAAATTAGA	2094	704	apneGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA	724
Db	2095	ATGAACAAACATTTGAGGCAGATGAGAA--	CTCCCATCA---	AATCAAAACAAGAGCTA	2148	724	snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyr	743
Qy	2149	TGAGAAATTTCTGGGATTTCTGAGTCTCTG---	TAGACTGTTTCCAGAGGATGGTG--	2203	743	GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy	763	
Db	2204	TTTACCAGGCTACACATCAAAAGAAATAGATAAAA--	TAATGGAAAAATAGAGAGTCT	2259	763	sLeuProLysAlaThrHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGluSe	783	
Qy	2260	TCT-GATAATGAGGTTTTCTGAAGGCTCCCT---	GAGAATGAAGTCTTATTCACATAAA	2315	783	TCT-GATAATGAGGTTTTCTGAAGGCTCCCT---	GAGAATGAAGTCTTATTCACATAAA	2315
Db	783	rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy	803	2316	GCCT--	TGATGATGCATGCATCAAACTTTCAAGAGCGGCTCCCG--	AGAGCCATCTG	2369
Qy	2316	GCCT--	TGATGATGCATGCATCAAACTTTCAAGAGCGGCTCCCG--	AGAGCCATCTG	2369	803	sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl	823
Db	2370	CTTCGAGCCTCGCATTTGAAAT---	GAAAAGTCTGTTCCAAATAAAC---	CTTGGAAATTAG	2423	823	aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys	842
Qy	2424	AATGAACACATTCAGACAGATCAGAT-----	TTCCCTTCAGATCAAAACAAAGAGG	2477	842	AATGAACACATTCAGACAGATCAGAT-----	TTCCCTTCAGATCAAAACAAAGAGG	2477
Db	843	AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys	861	2478	TTGAGAAAA--	TTCTGGGATTTCTGAAGTCT---	CCGTGGACTGTTTCACAGAGGATGG	2531
Qy	2478	TTGAGAAAA--	TTCTGGGATTTCTGAAGTCT---	CCGTGGACTGTTTCACAGAGGATGG	2531			


```
Db      862 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 881
QY      2532 TGTGTACC---CAGGCTACACACAAAGAAAGAGTAAAT---AGTGAGAAATTAGAAG 2585
Db      881 aLysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGluA 901
QY      2586 ATTCACTACCCATCAAAAT---CTTGTCAGTTCAT---CTTGTCAGTTCAT---CTTGTCAGTTCAT 2639
Db      901 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGlnArgAlaArgGluL 921
QY      2640 CTTCAAAAAGACACTGTGAACA-CGTACAGGAAATGGAACAAT---GAAAAAGAGTTT 2695
Db      921 euGlnLysAspHisCysGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 941
QY      2696 GTGTACTGAAAGAAC---TCTCAGAACAAAGAAATAAATCACAGTAGAGAA---CCA 2749
Db      941 ysValLeuLysLysLysLeuSerGluAlaLysGluLysSerGln-LeuGluAsnGln 960
QY      2750 AAAGTTAAATGGAAACAAGAG---TCTGCAGTGTAGTACTTAACCAA---GAAAG 2803
Db      961 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 980
QY      2804 AGAAGAGAGAAATGCCGATATATTAAAGAAAAA---TTAGGAAGAATTAG 2851
Db      981 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1000
QY      2852 AAGATCGAGACGACGATAGAAAGAGTTAGAGTGAACACACACTTGAAG---GCTCTC 2908
Db      1000 luGluGlnHisArg-----LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1017
QY      2909 AGATCAAG---ATAAGAAATCGAAGCTGTAGAAAGTAAATTTCAATAGCTTTCTCA---AC 2962
Db      1018 ArgIleGlnAspIleGluLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1037
QY      2963 TCATGAATGAAA---ATTATCTTACATGAATTTGCATGTGTGAAAGGAATGCCA--- 3017
Db      1037 rHisGluAsnGluAsnTyrLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1057
QY      3018 -TGCAAACTGGAATAGCCACACTA---AACACCAATCCAGAAAGGAAATAATC-T 3072
Db      1057 tleuLysLeuGluIleAlaThrLeuLysHis-GlnTyr-GlnGluLysGluAsnLysTyrP 1077
QY      3073 TTGAGACTTAAGATTTAAAGA---AAAGATGCTGAACCTTAG-ATGACCTTAAACTGA 3128
Db      1077 heGluAspIleLysLysLeuLysGluLysAsnAlaLeuGlnMetThrLeuLysLeuL 1097
QY      3129 AAAGGAATCATTAATAAG---GGATCTCAATATGTGGCAGCT---AAAGTTCTGATA 3182
Db      1097 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuL 1116
QY      3183 GCTGAGAAACAATGCTCATCTTAAATTTGAGGAAACACAGACAAAGAAAA---CTAGA 3236
Db      1117 AlaGlu-AsnThrMetLeuThrSerLysLysLysGluLysGlnAspLysGluIleLeuG 1136
QY      3237 GCAGAAATTTGAATA-CACCATCTGACTGGCTTCGCTGTACAGC-----CA 3283
Db      1136 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1153
QY      3284 TGATCAATTTGACATCATCAAGAAAGTAAGA---ACCTGCTTCCACATTTAG---GAGAT 3337
Db      1153 shspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1173
QY      3338 GCTGTTTCAAGAAATGAATGTTGATGTG---AGAGTACGATATTAACATGAGTGCT 3394
Db      1173 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluVal 1193
QY      3395 CCATCACCCTTTC---TGAGCTCAAGGAAATCCAAACCTTAAATAATTTCT---CAAT 3448
Db      1193 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsn 1213
QY      3449 TAGCAGGAGATGCTTAAGAGA---AATCATTTGTTTCAGAACATCACAAGAGA---CA 3502
```

```
1213 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1232
3503 ACCTGAAAAACAGTGTCAAT---GAAGGAAGCGACACATGTATCAAAAC-----GACAA 3553
1232 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyrGlnAsnGluGlnAspAs 1252
3554 GATAATGGACACAAACCTGACAGCAGAGTCTCTAGTCAGAAATTTATTTCAACTACAAG 3613
1252 nValAsnLysHisThrGluGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlns 1272
3614 CAAAAATTTGGTCTCACAGCAATT---AGTCATGCACATAGAAAGCTGACAAACAAAG 3669
1272 erLysAsnMetTyrLeuGlnGlnLeuValHisAlaHisLys-----LysA 1288
3670 CAGATAACAAT-----GATATTCATTTCTTG---AGAGGAATGCAACAC 3711
1288 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1308
3712 ATCTCTAAAGAGAAAAAGAGAGATATT---AATTCAATACCATTTTAAAAAACCCTAT 3768
1308 isLeuLeuLysGluLysAsnGlu-GluIlePheAsnTyrAsnAsnHisLeuLysAsnArg 1327
3769 ATT---CAATATGAAAAAGAGAAA 3789
1328 IleTyrGlnTyrGluLysGluLys 1335

RESULT 6
US-10-007-805-573
; Sequence 573, Application US/10007805
; Publication No. US20020150581A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yugu
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: McNeill, Patricia D.
; APPLICANT: Durham, Margarita
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C10
; CURRENT APPLICATION NUMBER: US/10/007,805
; CURRENT FILING DATE: 2001-12-07
; NUMBER OF SEQ ID NOS: 593
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 573
; LENGTH: 1349
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-007-805-573

Alignment Scores:
Pred. No.: 1,03e-108 Length: 1349
Score: 1223.00 Matches: 677
Percent Similarity: 59.07% Conservative: 150
Best Local Similarity: 48.36% Mismatches: 380
Query Match: 17.26% Indels: 207
DB: 13 Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-007-805-573 (1-1349)
QY 96 CAAAGGAAGACAGACATCACTTAATATACAGAG---CCGACAGAGACTGCTCTAAC 152
Db 11 LysArgLysLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu-- 29
QY 153 TGGCGCTGTGTCATGSCCT-----GAGGAAGTAGTAACATTTTC---GGTA 194
Db 30 -----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuValA 46
```

QY 195 GACAGAGTCCAGCTGACGCTCTTGG---TGCGAACAACGACACCTCTGATGAAGGC-T 250
Db : : : : :
46 spA1gYsCysGlnProAspValLeuAspGlyGluHisArgThrProLeuMetIysAlaL 66
QY 251 TACATGCCACAGAGGCTTGGC---AAATATTGATAGATTC-GGTGCCGATATAATC 306
Db : : : : :
66 euGInCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL 86
QY 307 TCGTG---ATGTGTATGCAACATGCTTCCATTATGCTTTATAGTGGATTTTTCAGT- 362
Db : : : : :
86 euValAspValTyGlyAsnMetAlaLeuHisTyAlaValTySerGluIleLeuSerV 106
QY 363 ---GGTGGCAAACTGCTGTC-CATGGTCAGATCGAAGTGC---CAACAAGCTGCTCA 416
Db : : : : :
106 alValAlaLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT 126
QY 417 CACCACCTTTTACTACCAT---AACGAAAGAGTCAAGCAATTGTGCAATTTT---GCTGAT 470
Db : : : : :
126 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuLe 145
QY 471 AAAAATGCAATGCAATGCTTA---ATAAGTTAAATGCACCCCTCATGCTGCTGAT 527
Db : : : : :
146 LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa 165
QY 528 G---TCTGGATCATCAGATAGTGTGATGCTTCTTAGCAAAA---TGTCAGCGTCTTT-- 579
Db : : : : :
165 lCysHisGlyLeuSerGluIleValGlyMetLeuLeuGlnGlnAsnValAspValPheAl 185
QY 580 -GTGCAGATATAGTGGATTAAGTCTGAGAACAT---ATGCTGTATCTGATTTCTCACA 635
Db : : : : :
185 alaAspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisI 205
QY 636 TTCATAACAAA---TTATGAATATATACGAAAATATCAAAATCA---TCAATATACAA 689
Db : : : : :
205 leHisGluGlnIleMet-GlutyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs 224
QY 690 T---CAGAGAACTCTGCAGAAACCTGTAGAGCTGCACCCCT---GGCGGAAAGACAC 743
Db : : : : :
224 nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP 244
QY 744 CTGACACGCTGAAGACT---GGTGGAAAACACCTGATAGGCTGCACCTT---GGTGA 797
Db : : : : :
244 roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu 263
QY 798 AAGACACCTGACAGCTGAAAGCT---GGTGGAAAACACCTGATGAGCTGATCT--- 851
Db : : : : :
264 ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu 283
QY 852 TGGTGGAGAACATCTGCACAAATTCATGTTGGAGAAGG---ACATCTCGAAAGTTTGA 908
Db : : : : :
284 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheG 303
QY 909 ACCTCAGC---AGAAAACACCTAGGAATATACATCC---TGCAAAAACACATCTGAGA 962
Db : : : : :
303 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL 323
QY 963 AATTTCAG---GGCCAGCAAGAGAGACCTAGAGATCGCAGGAGAGAA---AAAGAAGAC 1016
Db : : : : :
323 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 342
QY 1017 ACACCTAGGAAATATGAGTCCGAAAAGAAACACTGAGAAA-----TTA 1061
Db : : : : :
343 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 362
QY 1062 CGTGGCAGAAAAGAGACCTAGGAACCCATGGGAGAAAAGAAA---CCCTGTAAA 1118
Db : : : : :
363 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 379
QY 1119 GATGATGCGT-GCAAGAGTAACTCTAATAAACTAAAG---TTTGAAGAAAGAAATCT 1174
Db : : : : :
380 ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 399
QY 1175 AAGATGTT-GCATGTCCACAAAG---AATCATCTACAGCAAGTCCCATGATCAGAG 1230

Db 400 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg 419
QY 1231 ---TCCCATCAGATCCAAACAAGGAAGATGAGATATTTCTTT-CATTCTCGGTCTCT 1286
Db : : : : :
420 PheProSerGluSerLysGlnGluGluAspGluGluTrpSerCysAspSerArg---Ser 438
QY 1287 TTG-----AAGTTCTGCAAGATTCAAGTG---GTATACCTGAGTCTATATATAAAAGT 1337
Db : : : : :
439 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyGlnLysVal 458
QY 1338 AATGAGATAAATAAGAG---TAGAAGAGCTCTCTAAAGCCATCTGCTTCAAGCGCCA 1394
Db : : : : :
459 MetGluIleAsnArgGluValGluGluProProLysLysProSerAlaPheLysProAla 478
QY 1395 TTGAAAGCAAAA---CTCTTTCAAATAAGCCTTTG-----AATCAAGAATGAACA 1442
Db : : : : :
479 IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln----- 496
QY 1443 AACATTGAAGCAGATCCGTG-TTCCACCG---AATCCAAAACAAGGACTATAAGAAAT 1498
Db : : : : :
497 ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyGluGluAsn 516
QY 1499 TCTTGGATTCTAGATCTCTG---GAGACTGTTTACAGAAGG---ATTGTGTTTACCAA 1552
Db : : : : :
517 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy 536
QY 1553 GGCTACACATCAAAAGAATAGATAAAA---AATGAAAA---TAGAAGATCTCCTAAT 1606
Db : : : : :
536 sAlaThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL 556
QY 1607 AAGTGGTCTTCTGAGGCTACTCGAAT---GAAATTTCTATTCCACTAAGCCT--- 1660
Db : : : : :
556 ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 575
QY 1661 AGAATTGAAGACATGCAAACTTTCAAGCGAGCCTCC---GGGAAGCCATCTGCT---TC 1714
Db : : : : :
576 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheG 595
QY 1715 GAGCTCCTCACTGAAATCAAAAGTCTGTCCCAATAGCCTTG-----AATGAAA 1765
Db : : : : :
595 uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu- 614
QY 1766 ATGAAAACATCGAGGCGACATGAGAA---CTCCCATCAGAAATCCAAACAAAGGACTATA 1822
Db : : : : :
615 ---GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyG 633
QY 1823 GAAAA---TTCTGGGATACTGGAGTCTCTG---TAGACTGTTTTCACAGAGATTTG-TGT 1875
Db : : : : :
633 luGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 652
QY 1876 TTACCAAGG---CTCGGCTCAAAAAGATAGATAAAATATGAAAATTTAGAGGCT--- 1929
Db : : : : :
653 LeuProLysAlaAlaHisGlnLysGluIleAspLysIleAsnGlyLysLeuGluGlySer 672
QY 1930 CCTGTGTAAGTGGTCTTCTGAGGCTAACT---GGGAATGAAAAGTTTCTATTCCAATAAA 1986
Db : : : : :
673 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 692
QY 1987 GCCTAGA---ATTGATGACATCGAACTTCAAG---CAGGCTCCCGAGAGCCATCTCC 2040
Db : : : : :
693 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 712
QY 2041 CTTTCAGGCTCCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGAATTAAGA 2094
Db : : : : :
712 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 732
QY 2095 ATGAACAAACATTGAGGCGAGATGAGAA---CTCCCATCA---AATCCAAACAAGAGACTA 2148
Db : : : : :
732 snGluGlnThrLeuArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTy 751
QY 2149 TGAGAAAGTCTTGGGATTTCTGAGTCTCTG---TAGACTGTTTCCAGAGGATGCTG--- 2203

Db 752 GluGluSerSerTrpAspSer-GluSerLeuGluThrValSerGlnLysAspValCy 771
Qy 2204 -TTTACCAGGCTACACATCAAAAGATAGATAAAA---TAATCGAAATAGAGAGTC 2259
Db 771 sLeuProLysAlaThrHisGlnLysGluLeuAspLysLeuGluLeuGluSe 791
Qy 2260 TCT-GATTAATGAGGTTTCTGAGGCTCCCT---GAGNATGAAAGTTCTATTCCACTAAA 2315
Db 791 rProAspAsnAspGlyPheLeuLysAlaProCysArgMetLysValSerIleProThrLy 811
Qy 2316 GCCT---TCGATTGATCGATCGCAAACTTCAAGAGCGCCCTCCG---AGAGCCATCTG 2369
Db 811 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProGluLysProSerAl 831
Qy 2370 CTTGAGCGCTGATTGAAAT---GAAAGTCTGTTCMAATAAAC---CTTGGAAATTAAG 2423
Db 831 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 850
Qy 2424 AATGACACACATTGAGACAGATCAGAT---TTCCCTTCAGAAATCAAAACCAAGAGG 2477
Db 851 AsnGluGlnThr---LeuArgAlaAspGlnMetPheProSerGluSerLysGlnLysLys 869
Qy 2478 TTGAGAAA---TTCTGGGATTCGAAGTCT---CCGTGGACTGTTTCACAGAGGATGG 2531
Db 870 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 889
Qy 2532 TGTGTACC---CAGCTACACAAAAGAAAGGATAAAAT---AGTGGAAATTAAG 2585
Db 889 aLysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu 909
Qy 2586 ATTCACTAGCCTATCAAAAT---CTTGTACAGTTTCAT---CTTGTGAAAGCGCAAGGAA 2639
Db 909 sPheSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaArgGlu 929
Qy 2640 CTTCAAAAAGACACTGTGAACA-GTACAGAAAATGGAACAAT---GAAAAGAGTTT 2695
Db 929 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysLysPheC 949
Qy 2696 GTGTACTGAAAGAAC---TGTCAGAAACAAAGAAATAAATCACAGTAGAGAA---CCA 2749
Db 949 ysValLeuLysLysLysLeuSerGluAlaLysGluLeuLysSerGln-LeuGluAsnGln 968
Qy 2750 AAAGTTAAATGGACAAAGAG---TCTCAGTGTAGATGCTTAAACCAA---GAAAAG 2803
Db 969 LysValLysTrpGluGlnGluLeuLysSerValArgLeuThrLeuAsnGlnGluGlu 988
Qy 2804 AGACAGAGAAATGCCGATATATTAAGAAAAA---TTAGGAGAAATAG 2851
Db 989 LysArg-ArgAsnAlaAspIleLeuAsnGluLysIleArgGluGluLeuGlyArgIle-G 1008
Qy 2852 AAGAAATCGAGACGATAGAAAGAGTTAGAAAGTGAACACACAACTTGAAG---GCTCTC 2908
Db 1008 luGluGlnHisArg---LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1025
Qy 2909 ACATACAG---ATGAAATGCAAGCTGTAGAAATGTAATGTAATGTTCTCA---AC 2962
Db 1026 ArgIleGlnAspIleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1045
Qy 2963 TCATCAAAATGAAA---ATTATTCTTACATGAAATGCTGTTGAAAAGGAATGCGCA-- 3017
Db 1045 rHisGluAsnGluAsnTyLeuLeuHisGluAsnCysMetLeuLysLysGluLeuAlaMe 1065
Qy 3018 -TGCAAAACTGGAATAGCCACACTA---AACACCAATCCAGAAAAGAAAATAATC-T 3072
Db 1065 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyGlnGluGlyGluAsnLysTyP 1085
Qy 3073 TTGAGACTTAAGATTTAAAGA---AAAGATGCTGAACTTAG-ATGACCCCTAAATCGA 3128
Db 1085 heGluAspIleLysLysLysGluLysAsnAlaGluLeuGlnMetThrLysLysLeuL 1105
Qy 3129 AAAGGAATCACTACTAAAG---GGATCTCAATATGTGGCAGCT---AAGTCTCTGATA 3182
Db 1105 ysGluGluSerLeuThrLysArgAlaSerGlnTySerGly-GlnLeuLysValLeuIle 1124

Qy 3183 GCTGAGAAACAATGCTCTATTCTAAATTTAGGAGAAAACACAGCAAGAAA---CTAGA 3236
Db 1125 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGlnAspLysGluIleLeuGl 1144
Qy 3237 GCAGAAAATTGAATA-CACCATCTGACTGGCTTCGCTGTACAAGC-----CA 3283
Db 1144 uAlaGluIleGluSerHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1161
Qy 3284 TGATCAATTTGACATCAAGAAAAGTAAGA---ACCTGCTTCCACATGTAG---GAGAT 3337
Db 1161 sAspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1181
Qy 3338 GCTGTTTGCAGAAAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACATGAGTCT 3394
Db 1181 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTy-AsnAsnGluValL 1201
Qy 3395 CCATCACCACCTTTC---TGAGCTCAAGGAAATCCAAAACCTTAAATATATCT---CAAT 3448
Db 1201 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLysLysIleAsnLeuAsnT 1221
Qy 3449 TAGCAGGAGATGCTTAAGAGA---AATACATTTGTTTCAGACATCATCACAAGAGA---CA 3502
Db 1221 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1240
Qy 3503 ACGTGAACACAGTGTCAAT---GAAGGAGCGAACACACATGTATCAAAAC-----GACAA 3553
Db 1240 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyGlnAsnGluGlnAspAs 1260
Qy 3554 GATAATGGAACAAACACCTGAAACAGCAGAGTCTCTAGTCAGAAATTTTCAACTACAAG 3613
Db 1260 nValAsnLysHisThrGluGlnGlnGlnSerLeu-AspGlnLysLeuPheGlnLeuGlnS 1280
Qy 3614 CAAAATTTGCTTCACAGCAAT---AGTCATGCAATAGAAAGCTGACAAACAAAAG 3669
Db 1280 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1296
Qy 3670 CAGATAACAAT-----GATATTCATTTCTTG---AGAGGAATGCAACAC 3711
Db 1296 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHisH 1316
Qy 3712 ATCTCTTAAAGAGAAAAGAGAGAGATATT---AATTACAATACCATTTAAAAACCGTAT 3768
Db 1316 isLeuLeuLysGlnLysAsnGlu-GluIlePheAsnTyAsnAsnHisLeuLysAsnArg 1335
Qy 3769 ATT---CAATATGAAAAGAGAAAAGAGAAACAGAAACTCATG 3807
Db 1336 ileTyGlnTyGlnLysGluLysAlaGluThrGluValIle 1349

RESULT 7

US-10-076-622-573
; Sequence 573, Application US/10076622
; Publication No. US20030023036A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.470C11
; CURRENT APPLICATION NUMBER: US/10/076,622
; CURRENT FILING DATE: 2002-02-13
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 573
; LENGTH: 1349
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-076-622-573

Alignment Scores: 1.03e-108 Length: 1349
Pec. No.: 1223.00 Matches: 677
Score:

Percent Similarity: 59.07% Conservative: 150
Best Local Similarity: 48.36% Mismatches: 380
Query Match: 17.26% Indels: 207
DB: 14 Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-076-622-573 (1-1349)

QY 96 CAAGGAGAGAGACATCAACTTAATATACAGAG--CCACAGAGAGACTGCTCTAAC 152
DB 11 LysArgLysThrIleasnLeuasnIleGlnaspAlaGlnlyeArg-ThrAlaLeu-- 29
QY 153 TGGGCGCTGCTCAATGGCCT-----CAGGAAGTAGTAACATTTCC---GGTA 194
DB 30 -----HisTrpAlaCysValasnGlyHisGluGluValValThr-PheLeuValA 46
QY 195 GACAGAAGTCCAGCTGAGCTCTG---TGCGCAACACGAGACCTCTGATGAAGC-T 250
DB 46 sPArgLysCysGlnProaspValLeuaspGlyGluHisArgThrProLeuMetLysAla 66
QY 251 TACAATGCCACAGGAGGCTTTGC-----AAATATTGATAGATTC-GGTCCCGATATAATC 306
DB 66 euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleaspSerGlyAlaaspIleAsnL 86
QY 307 TCGTG--ATGTGATGCAACATGCTTCCATTATGCGTTTATAGTGATTTTGTAGT- 362
DB 86 euValAspValTyroGlyAsnMetAlaLeuHisTyAlaValTy-SerGluIleLeuSerV 106
QY 363 --GGTGGCAAACTGCTGTC-CATGTGCGAGTATCGAAGTGC---CAACAAGGCTGCCTCA 416
DB 106 alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisasnLysAlaSerLeuT 126
QY 417 CACCACTTTTACTACCAT---AACGAAAGAGTGAAGCAATTTGTGGAATTT--GCTGAT 470
DB 126 hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuIle 145
QY 471 AAAAATGCAATGCAATGCGTTA--ATAAGTTAAATGCACACCTCATGCTCTGTAT 527
DB 146 LysasnAlaAsnAlaasnAlaValasnLys-TyryLysCysThrAlaLeuMetLeuAlaVa 165
QY 528 G---TCTGGATCATCAGAGATAGTTGTCATGCTCTTACGAAA---TGTGACGCTCTTT-- 579
DB 165 lCysHisGlyLeuSerGluIleValGlyMetLeuLeuGlnGlnAsnValaspValPheAl 185
QY 580 -GTCAGATATAGGAGTACTCAGACAT---ATGCTGTACTGTGATTTCTCACA 635
DB 185 alaaspIle-CysGlyValThrAlaGluHisTyAlaValThrCysGlyPheHisI 205
QY 636 TTCATAACAAA---TTATGAATATATACGAAATATCAAAATCA---TCAAAATACAA 699
DB 205 leHisGluGlnIleMet-GluTyriLeargLysLeuSer-LysasnHisGlnasnThrAs 224
QY 690 T---CAGNAGGAACCTGCAGAAACCTGATGAGGCTGCACCTT---GGCGGAAAGACAC 743
DB 224 nProGluGlyThrSerAlaGlyThrProaspGluAlaAla-ProLeuAlaGluArgThrP 244
QY 744 CTGACACGCTGAAAGCT---GGTGGAAAAACACCTGATGAGGCTGCACCTT---GGTGA 797
DB 244 roaspThrAlaGluSerLeuValGlnLysThrProaspGluAlaAla-ProLeuValGlu 263
QY 798 AAGACACTGACAGGCTGAAGCT---GGTGGAAAAACACCTGATGAGGCTGCATCT--- 851
DB 264 ArgThrProaspThrAlaGluSerLeuValGlnLysThrProaspGluAlaAlaSerLeu 283
QY 852 TGGTGGAGGAACATCTGACAAATTCATGTTGGAGAAAGG---ACATCTGGAAGTTGGA 908
DB 284 Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl 303
QY 909 ACCTGACG---AGAAAAACACCTAGGAATATACATCC---TGCAAAAAACATCTGAGA 962
DB 303 uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThr-SerGluL 323
QY 963 AATTAGG---GGCCCAAGAGACCTAGAGATCGCAGGAGAA---AAAGAAGAC 1016
DB 963 ProValLysAspGlyLeuLeuLysAlaasnCysGlyMetLysValSerIleProThrLys 692

DB 323 ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp 342
QY 1017 ACACCTAGGGAATATCAGTCCCGAAAGAAACAGACTCAGAAA-----TTA 1061
DB 343 ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla 362
QY 1062 CGTGGGAGAAAAGGAGACACTAGGAAGACGATGGAGAGAAAAGAAA---CCCTGTAAA 1118
DB 363 LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys 379
QY 1119 GATGGATGCGT-GCAAGAGTAACATCTATAATAACTAAAG--TTTTGAAAAGGAATCT 1174
DB 380 ThrGlyCysValAlaargValThrSerAsnLysThrLysValLeuGluLysGlyArgSer 399
QY 1175 AAGATGTT-GCATGTCACAAAAG---AATCATCTACAAAGCAAGTGCATGATCAGAG 1230
DB 400 LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaasnAspGlnArg 419
QY 1231 ---TCCCATCAGATCCAAACAAGGGAAGATGAAGAATATCTTT-CATTCTCGGCTCTCT 1286
DB 420 PheProSerGluSerLysGlnGluAspGluLutyrSerCysaspSerArg---Ser 438
QY 1287 TTG-----AAGTCTGCAAGATTCAAGTG---GTATACCTCAGTCTATATATAAAAGT 1337
DB 439 LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleTyroGlnLysVal 458
QY 1338 AATGAGATAAATAAGAG---TAGAAGGCTCTAAAGCCCATCTCCCTTCAGCCGCCA 1394
DB 459 MetGluIleasnArgGluValGluGluProLysLysProSerAlaPheLysProAla 478
QY 1395 TTGAAAGCAAAA---CTCTTTCCAAATAAGCCCTTTG-----AATGAAAGATGAACA 1442
DB 479 IleGluMetGlnAsnSerValProasnLysAlaPheGluLeuLysasnGluGln----- 496
QY 1443 AACATTGAAGCAGATCCGTG-TTCCACCG---AATCCAAACAAGGACTATAAGAAAT 1498
DB 497 ThrLeuArgAlaaspProMetPheProGluSerLysGlnLysAspTyroGluGluasn 516
QY 1499 TCTTGGGATTCAGAGTCTCTG---GAGACTGTTTACAGAAG---ATTGTGTTTACCAA 1552
DB 517 SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProly 536
QY 1553 GGTACACATCAAAAGAATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT 1606
DB 536 salThrHisGlnLysGlu-IleAspLysIleasnGlyLysLeuGluSerProAsnL 556
QY 1607 AAAGTGTCTCTGAGGCTACTCGGAAT---GAAATTTCTATCCACTAAGCCT--- 1660
DB 556 ysaspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu 575
QY 1661 AGAATTGAAGACATGCAAACTTTCAAAGCGAGCCCTCC---GGGAAGCCATCTGCT---TC 1714
DB 576 GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl 595
QY 1715 GAGCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTG-----AATGAAA 1765
DB 595 uProAlaThrGluMetGlnLysSerValProasnLysAlaLeuGluLeuLysasnGlu-- 614
QY 1766 ATGAAAAACATCGAGCGCAGATGAGAA---CTCCCATCAGAATCCAAACAAGGACTATA 1822
DB 615 ---GlnThrTrpArgAlaAsp-GluIleLeuProSerGluSerLysGlnLysAspTyroG 633
QY 1823 GAAAA---TTCGGGATGACTGAGTCTCTG---TAGACTGTTTACAGAAGATTTG-TGT 1875
DB 633 luGluasnSerTrpaspThr-GluSerLeuCysGluThrValSerGlnLysAspValCys 652
QY 1876 TTACCAAGG---CTCGCTCAAAAGAATAGATAAAATAATGGAATAATAGAGGT--- 1929
DB 653 LeuProLysAlaAlaHisGlnLysGluIleaspLysIleasnGlyLysLeuGluGlySer 672
QY 1930 CCTGTTAAAGTGTCTTCTCAGGCTAACT---GGGAATGAAAGTCTTATTTCCAACTAAA 1986
DB 673 ProValLysAspGlyLeuLeuLysAlaasnCysGlyMetLysValSerIleProThrLys 692

; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.470C12
; CURRENT APPLICATION NUMBER: US/10/124,805
; CURRENT FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 627
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 573
; LENGTH: 1349
; TYPE: PRN
; ORGANISM: Homo sapiens
; US-10-124-805-573

Alignment Scores:
Pred. No.: 1.03e-108 Length: 1349
Score: 1223.00 Matches: 577
Percent Similarity: 59.07% Conservative: 150
Best Local Similarity: 48.36% Mismatches: 207
Query Match: 17.26% Indels: 207
DB: 14 Gaps: 115

US-09-602-362E-22 (1-4115) x US-10-124-805-573 (1-1349)

QY	96	CAAGGGAAGAGACATCAACTTAATATACAGAG-----CCGAGAGAGACTGCTTAAC	152
DB	11	LysArgLysThrIleAsnLeuAsnIleGlnAspAlaGlnLysArg-ThrAlaLeu--	29
QY	153	TGGGCTGGTCAATGGCCT-----GAGGAAGTACTAATTC---GGTA	194
DB	30	-----HisTrpAlaCysValAsnGlyHisGluGluValValThr-PheLeuVala	46
QY	195	GACAGAGTCCAGCTGAGTCCCTTG---TGGCGAACACGGACACCTCTGATGAAGCC-T	250
DB	46	spA:rgLysCysGlnProAspValLeuAspGlyGluHisArgThrProLeuMetLysAlaL	66
QY	251	TACAAATGCCAGAGGCTTGGC---AAATATTTGATAGATTC-GTGCGCGATATAATC	306
DB	66	euGlnCysHisGlnGluAlaCysAlaAsnIleLeuIleAspSerGlyAlaAspIleAsnL	86
QY	307	TCGTG---AUGTGTATGCAACATCGCTTCCATTATGCGTTTATAGTGGATTTTGTCAGT-	362
DB	86	euValAspValTyrGlyAsnMetAlaLeuHisTyrAlaValTyrSerGluLeuLeuSerV	106
QY	363	--GGTGGCAACTGCTCTC-CATGGTCAGTATCGAAGTGC---CAACAAGGCTGCTCA	416
DB	106	alValAlaLysLeuLeuSerHisGlyAlaValIleGluValHisAsnLysAlaSerLeuT	126
QY	417	CACCCTTTTACTACCAT---AACGAAAGAGTGAACATTTGTGAATTTT---GCTGAT	470
DB	126	hrProLeuLeuLeu-SerIleThrLysArgSerGluGlnIleValGluPheLeuLeuIle	145
QY	471	AAATATCAATCGAATGCTTA---ATAAGTTAATGCACACCTCATGCTGCTGAT	527
DB	146	LysAsnAlaAsnAlaAsnAlaValAsnLys-TyrLysCysThrAlaLeuMetLeuAlaVa	165
QY	528	G---TCTGGATCATCAGAGATAGTTGCATGCTTCTTAGCAAAA---TGTGACGCTCTTT--	579
DB	165	lCysHisGlyLeuSerGluLeuValGlyMetLeuLeuGlnGlnAsnValAspValPheal	185
QY	580	-GTGAGATATAGTGGAGTAACTCAGAACAT---ATGCTGTACTGTGGATTCTTCACA	635
DB	185	aAlaAspIle-CysGlyValThrAlaGluHisTyrAlaValThrCysGlyPheHisIst	205
QY	636	TTCAATAACAA---TTATGAATATATACGAATATATCAAAATCA---TCAAAATACAA	689
DB	205	leHisGluGlnIleMet-GluTyrIleArgLysLeuSer-LysAsnHisGlnAsnThrAs	224
QY	690	T---CAGAAGAACTTCGAGGAAACCTGATGAGGTGCACCTC---GGCGGAAAGACAC	743
DB	224	nProGluGlyThrSerAlaGlyThrProAspGluAlaAla-ProLeuAlaGluArgThrP	244
QY	744	CTGACGCTGAAGCT---GGTGGAAACACCTGATGAGGTGCACCTC---GGTGA	797
DB	244	roAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAla-ProLeuValGlu	263

QY	798	AAGACACCTGACAGGCTGAAAGCT---GGTGGAAACACCTGATGAGGTGCTATCT---	851
DB	264	ArgThrProAspThrAlaGluSerLeuValGluLysThrProAspGluAlaAlaSerLeu	283
QY	852	TGGTGGAGGAACATCTGACAAATTCATGTTGGAGAAAGG---ACATCTGGAAGTTGGA	908
DB	284	Val-GluGlyThrSerAspLysIleGlnCysLeuGluLysAlaThrSerGlyLysPheGl	303
QY	909	ACGTCAGC---AGAAAAACACCTAGGAATTTACGATCC---TGCAAAAACACCTCTGAG	962
DB	303	uGln-SerAlaGluGluThrProArgGluIleThrSerProAlaLysGluThrSerGluL	323
QY	963	AATTTAGC---GGCCAGCAAGGAACCTAGAGATCCAGGAGAA---AAAGAAGAC	1016
DB	323	ysPheThrTrpPro-AlaLysGlyArgProArgLysIleAlaTrpGluLysLysGluAsp	342
QY	1017	ACACCTAGGGAATATGAGTCCCGAAAAAGAACACCTGAGAA-----TTA	1061
DB	343	ThrProArgGluIleMetSerProAlaLysGluThrSerGluLysPheThrTrpAlaAla	362
QY	1062	CTGGGCGAGAAAGGAGACCTAGGAAGACGCTGGAGAGAAAGAAA---CCCTGTAAA	1118
DB	363	LysGlyArgProArgLysIle-----AlaTrpGluLysLysGluThrProValLys	379
QY	1119	GATGGATCGCT-CCAGAGTACATCTAATAACTAAAG---TTTTGAAAAAGGAATCT	1174
DB	380	ThrGlyCysValAlaArgValThrSerAsnLysThrLysValLeuGluLysGlyArgSer	399
QY	1175	AAGATGTT-GCATGTCCACAAAAG---AATCATCTACAAAGCAAGTGCCTCATGACAGG	1230
DB	400	LysMetIleAlaCysProThrLysGluSerSerThrLysAlaSerAlaAsnAspGlnArg	419
QY	1231	---TCCCATCAGATCCAAACAGGAGATGAAGATATCTTT-GATTCTCGGCTCTCT	1286
DB	420	PheProSerGluSerLysGlnGluAspGluGluTyrSerCysAspSerArg---Ser	438
QY	1287	TTTG-----AAGTCTTCGCAAGATTCAAGTG---GTATACCTCAGTCTATATATAAAGT	1337
DB	439	LeuPheGluSerSerAlaLysIleGlnValCysIleProGluSerIleThrGlnLysVal	458
QY	1338	ANTGAGATAAATAGAAAG---TAGAAAGCCTCCTAAAAGCCTCTGCCTTCAAGCGCCA	1394
DB	459	MetGluIleAsnArgGluValGluGluProLysLysProSerAlaPheLysProAla	478
QY	1395	TTGAAAGCAAAA---CTCTTTCCAAATAGCCTTTG-----AATGAAGAAATGAACA	1442
DB	479	IleGluMetGlnAsnSerValProAsnLysAlaPheGluLeuLysAsnGluGln-----	496
QY	1443	AACATTGAAGCAGATCCGCTG-TTCCCACCG---AATCCAAACAAAGGACTATAAGAAAAT	1498
DB	497	ThrLeuArgAlaAspProMetPheProGluSerLysGlnLysAspTyrGluGluAsn	516
QY	1499	TCTTGGGATCTAGAGTCTCTG---GAGACTGTTTACAGAGG---ATTGTGTTTACCAA	1552
DB	517	SerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCysLeuProLy	536
QY	1553	GGCTACACATCAAAAGAAATAGATAAAA---AATGGAAAA---TAGAAGAGTCCCTAAT	1606
DB	536	salalThrHisGlnLysGlu-IleAspLysIleAsnGlyLysLeuGluGluSerProAsnL	556
QY	1607	AAAGTGGTCTTCTGAGGCTACCTCGGAAT---GAAATTTCTATTCCACTAAAGCCT---	1660
DB	556	ysAspGlyLeuLeuLysAlaThrCys-GlyMetLysValSerIleProThrLysAlaLeu	575
QY	1661	AGAATTGAAGACATCCAAACCTTTCAAGCGAGCCTCC---GGGAAGCCATCTGCT---	1714
DB	576	GluLeuLysAspMetGlnThrPheLysAlaGlu-ProProGlyLysProSerAlaPheGl	595
QY	1715	GAGCCTCCACTGAAATCAAAAGTCTGTCCCAATTAAGCCTTTGG-----AATGAAAA	1765
DB	595	uProAlaThrGluMetGlnLysSerValProAsnLysAlaLeuGluLeuLysAsnGlu--	614

1766 ATGAATAACATGGAGCAGATGAGAA---CTCCATCAGAAATCCAAACAAAGGACTATAA 1822
Db : : : : :
615 ----GlnThrTrpArgAlaasp-GluLeuLeuProSerGluSerGluLysGluLysAspTyrG 633
Qy : : : : :
1823 GAAAA---TTCTGGGACTGGAGTCTCTG---TAGACTGTTTTCAGAGGAGTGTG-TGT 1875
Qy : : : : :
633 LuGluAsnSerTrpAspThr-GluSerLeuCysGluThrValSer-GlnLysAspValCys 652
Db : : : : :
1876 TTACCAAGG---CTGGCTCATAAAGCAATAGATAAAATAATGAGAAATTAGAAGGGT--- 1929
Qy : : : : :
653 LeuProLysAlaAlaHisGlnLysGluLeuAspLysLeuAsnGlyLysLeuGluGlySer 672
Db : : : : :
1930 CCTGTGTTAAAGTGGTCTCTGAGCTAACT---GGGAATGAAGTTCTATTCCAACTAAA 1986
Qy : : : : :
673 ProValLysAspGlyLeuLeuLysAlaAsnCysGlyMetLysValSerIleProThrLys 692
Db : : : : :
1987 GCCTAGA---ATTGATGACATGCAPAACTTCAAG---CAGGCTCCCGAGGAGCCATCTGC 2040
Qy : : : : :
693 Ala-LeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 712
Db : : : : :
2041 CTTCGAGCCCTCCATTGAAA---TCAAAAGTCTGTCCAAATAAAC---CTTGGAATTAAGA 2094
Qy : : : : :
712 aPheGluProAlaIleGluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLysA 732
Db : : : : :
2095 ATGAACAAACATTTAGGAGCAGATGAGAA---CTCCCATCA---AATCCAAACAAAGGACTA 2148
Qy : : : : :
732 snGluGlnThrLeuAlaasp-GluLeuLeuProSerGluSerLysGlnLysAspTyr 751
Db : : : : :
2149 TGAGAAAGTTCTTGGGATCTGAGTCTCTG---TAGACTGTTTCCAGAGGAGTGTG--- 2203
Qy : : : : :
752 GluGluSerSerTrpAspSer-GluSerLeuCysGluThrValSerGlnLysAspValCy 771
Db : : : : :
2204 -TTTACCAGGCTACATCAACAAAGCAATAGATAAA---TAATGGAAATAGAGAGTCT 2259
Qy : : : : :
771 sLeuProLysAlaThrHisGlnLysGluLeuAspLysLeuGluLeuGluGluLe 791
Db : : : : :
2260 TCT-GATAATGAGGTTTCTGAGGCTCCCT---GAGAAATGAAGTTCTATTCCCACTAAA 2315
Qy : : : : :
791 rProAspAsnArgGlyPheLeuLysAlaProCysArgMetLysValSerIleP-oThrLy 811
Db : : : : :
2316 GCCT---TGAAATGATGGATGCAACTTTCAAGCAGGCTCCCG---AGAGCCATCTG 2369
Qy : : : : :
811 sAlaLeuGluLeuMetAspMetGlnThrPheLysAlaGluProProGluLysProSerAl 831
Db : : : : :
2370 CTTCGAGCCCTGCATGAAT---GAAAGTCTGTTCCAAATAAAC---CTTGGAAATTAAG 2423
Qy : : : : :
831 aPheGluProAlaIle-GluMetGlnLysSerValProAsn-LysAlaLeuGluLeuLys 850
Db : : : : :
2424 AATGACAAACATTTAGACAGATCAGAT---TTCCCTTCAGATCAAAACAAAGAGG 2477
Qy : : : : :
851 AsnGluGlnThr---LeuArgAlaaspGlnMetPheProSerGluSerLysGlnLysLys 869
Db : : : : :
2478 TTGAGAAA---TTCTGGGATCTGAGTCT---CCGTGGAGTCTTTCACAGAGGATGG 2531
Qy : : : : :
870 ValGluGluAsnSerTrpAspSerGlu-SerLeuArgGluThrValSerGlnLysAsp-V 889
Db : : : : :
2532 TGTGTACC---CAGGCTACACAAAGAGAGGATAAAAT---AGTGGAAATTAAG 2585
Qy : : : : :
889 alCysValProLysAlaThrHisGlnLysGluMetAspLysIleSerGlyLysLeuGlu 909
Db : : : : :
2586 ATTCACTAGCCTATCAAAAT---CTTGGTACAGTTTCAT---CTTGTGAAGCCAGGGA 2639
Qy : : : : :
909 spSerThrSerLeuSerLysIleLeuAspThrValHisSerCysGluArgAlaargGluL 929
Db : : : : :
2640 CTTCAAAAACACACTGTGAGCA-CGTACAGGAAAATGGAACAAT---GAAAAAGAGTTT 2695
Qy : : : : :
929 euGlnLysAspHisCysGluGlnArgThrGlyLysMetGluGlnMetLysLysPheC 949
Db : : : : :
2696 GTGTACTGAAAAGAAC---TTCTAGAACAAAGAAATAATATCAGTACAGTACAG 2749
Qy : : : : :
949 ysValLeuLysLysLysLeuSerGluAlaLysGluIleLysSerGln-LeuGluAsnGln 968
Db : : : : :
2750 AAAGTTAAATGGGAACAAGAG---TCTGAGTGTAGATTGACTTAAACCAA---GAAAAG 2803
Qy : : : : :

969 LysValLysTrpGluGlnGluLeuCysSerValArgLeuThrLeuAsnGlnGluGlu 988
Db : : : : :
2804 AGAAGAGAGAAATGCCGATATATTAAGAAAAA---TTAGGAAGAAATTAG 2851
Qy : : : : :
989 LysArg-ArgAsnAlaaspLleLeuAsnGluLysIleArgGluLeuGlyArgLle-G 1008
Db : : : : :
2852 AAGATCCGAGACGACATAGAAAGAGTGTAGAAAGTGAACACAACTTGAAG---GCTCTC 2908
Qy : : : : :
1008 LuGluGlnHisArg---LysGluLeuGluValLysGlnGlnLeu-GluGlnAlaLeu 1025
Db : : : : :
2909 AGATACAAG---ATAGAATTCAGGTGTGAAGTAATTTGAATAGATTCTCA---AC 2962
Qy : : : : :
1026 ArgLleGlnAspLleGluLeuLysSerValGluSerAsnLeuAsnGln-ValSerHisTh 1045
Db : : : : :
2963 TCATGAATGAAA---ATTATTCTTACATGAAATTCATGTTGAAAAGGAATTCGCA-- 3017
Qy : : : : :
1045 rhisGluAsnGluAsnTyrLeuLeuHisGluAsnCysMetLeuLysLysGluIleAlaMe 1065
Db : : : : :
3018 -TGCAAAACCTGGAATGAGCACACTA---AACACCAATCCAGAAAAAGAAAATAATC-T 3072
Qy : : : : :
1065 tLeuLysLeuGluIleAlaThrLeuLysHis-GlnTyrGlnGluLysGluAsnLysTyrP 1085
Db : : : : :
3073 TTGAGGACTTAAGATTTTAAAGA---AAGATGCTGAACTTAG-ATGACCCCTAAACTGA 3128
Qy : : : : :
1085 heGluAspLysLysLysLysLysGluLysAsnAlaGluLeuGlnMetThrLeuLysLeuL 1105
Db : : : : :
3129 AAAGGAATCATTTCTAAAG---GGATCTCAATATGTGGCGAGCT---AAGTTCTGATA 3182
Qy : : : : :
1105 ysGluGluSerLeuThrLysArgAlaSerGlnTyrSerGly-GlnLeuLysValLeuIle 1124
Db : : : : :
3183 GCTGAGAAACAAATGCTCATTTCTAAATTCAGGAAAAACAGACAAACAAA---CTAGA 3236
Qy : : : : :
1125 AlaGlu-AsnThrMetLeuThrSerLysLeuLysGluLysGlnAspLysGluIleLeuGl 1144
Db : : : : :
3237 GGCAGAAATTTGATA-CAACATCTCTGAGTGGCTTCGCTGTACAGC-----CA 3283
Qy : : : : :
1144 uAlaGluIleGluSerHisHisPro-----ArgLeuAlaSerAlaValGlnAspHi 1161
Db : : : : :
3284 TGATCAATTTGTGCATCAAGAAAGTAAGA---ACCTGCTTCCACATTCAG---GAGAT 3337
Qy : : : : :
1161 saspGlnIleValThrSerArgLysSerGlnGluProAlaPheHisIleAlaGlyAspAl 1181
Db : : : : :
3338 GCTGTTTCCAAAGAAATGAATGTTGATGTG---AGAGTACGATATTAAACAATGAGTGT 3394
Qy : : : : :
1181 aCysLeuGlnArgLysMetAsnValAspValSerSerThrIleTyr-AsnAsnGluValL 1201
Db : : : : :
3395 CCATCACCACCTTTC---TGAGCTCAAGGAAATCCAAACCTCAAAATTTATCT---CAAT 3448
Qy : : : : :
1201 euHisGlnProLeuSerGluAlaGlnArgLysSerLysSerLeuLysIleAsnLeuAsnT 1221
Db : : : : :
3449 TAGCAGGAGATGCTTAAGAGA---AATACATGTTGTTTCAGAACTCAACAAGAGA---CA 3502
Qy : : : : :
1221 yrAlaGlyAsp-AlaLeuArgGluAsnThrLeuValSerGluHis-AlaGlnArgAspGl 1240
Db : : : : :
3503 AGCTCAAAACACAGTGTCAAT---GAAGGAGGCAACACATGTATCAAAAC-----GACAA 3553
Qy : : : : :
1240 nArgGluThrGlnCysGlnMetLysGluAlaGluHisMetTyr-GlnAsnGluGlnAspAs 1260
Db : : : : :
3554 GATAATGGAACAAACACCTGAAACAGCAGTCTCTAGTACAGAAATATTTCACACTACAG 3613
Qy : : : : :
1260 nValAsnLysHisThrGluGlnGlnGluSerLeu-AspGlnLysLeuPheGlnLeuGlns 1280
Db : : : : :
3614 CAAAAATTTGTGCTTCACAGCAATT---AGTCATGCACATAGAAAGCTCAACAACAAAG 3669
Qy : : : : :
1280 erLysAsnMetTrpLeuGlnGlnGlnLeuValHisAlaHisLys-----LysA 1296
Db : : : : :
3670 CAGATAACAAT-----CATATTCTTTCTTG---AGAGGAATTCACACAC 3711
Qy : : : : :
1296 laAspAsnLysSerLysIleThrIleAspIleHisPheLeuGluArgLysMetGlnHis 1316
Db : : : : :
3712 ATCTCTCTTAAAGAGAAAAAGAGGAGATATT---AATTACAAATACCATTTTAAACCGTAT 3768
Qy : : : : :